Variational Objecthood: A Least-Complexity Principle for Stable Identity, Compression, and Concept Formation

Douglas C. Youvan

doug@youvan.com

www.youvan.ai

July 10, 2026

The recent algebra of object formation explains when an observation policy produces coherent quotient objects, when those objects remain stable under transformation, and when local identity systems descend faithfully into a global ontology. Yet the policy itself has remained externally supplied. Mathematics can test whether a proposed identity is admissible, but it does not yet determine which admissible identity policy should be chosen. This paper introduces a variational theory of objecthood. An identity policy is treated as an equivalence relation on a state space and evaluated by the compression it achieves, the task-relevant distortion it introduces, its invariance under admissible transformations, and its coherence under local-to-global closure. The preferred object system is the least complex quotient that preserves the distinctions required for prediction, explanation, intervention, or lawful dynamics. This converts object formation from a passive quotient construction into an optimization problem over the lattice of equivalence relations. The resulting principle provides a mathematical basis for discovering rather than merely assuming objects. It also suggests a general mechanism for abstraction, emergence, representation learning, scientific classification, and artificial intelligence: stable objects are optimal identity policies under preservation constraints.

Keywords: variational objecthood, object formation, concept formation, identity policy, equivalence relations, quotient spaces, stable quotients, conceptual resolution, admissibility, compression, rate-distortion, invariant representation, descent theory, observation policies, artificial intelligence, emergence, abstraction, scientific discovery, optimization over partitions.

A collaboration with OpenAI GPT-5.5 Thinking. CC4.0. 283 pages and 60 references.

https://orcid.org/0009-0004-4206-4253

Central Equation

Let X be a state space, Eq(X) its lattice of equivalence relations, G a family of admissible transformations, π a task or explanatory policy, and $C\mathcal{U}$ the local-to-global closure operator associated with an observational cover $\mathcal{U}$.

Define:

Dπ(E) = task-relevant distortion introduced by quotienting under E,

K(E) = complexity of the quotient object system X/E.

The preferred identity policy is

E★ ∈ arg min {Dπ(E) + λK(E) : E ∈ Eq(X)}

subject to the closure-coherence condition

$C\mathcal{U}$(E) = E

and the transformation-invariance condition

x E y implies g(x) E g(y)

for every g ∈ G.

Equivalently, the preferred policy minimizes

least relevant distortion + weighted complexity

within the class of identity systems that are both closure-coherent and transformation-invariant.

The complete constrained form is

E★ ∈ arg min {Dπ(E) + λK(E) : E ∈ Fix($C\mathcal{U}$) ∩ Inv(G)},

where

Fix($C\mathcal{U}$) = {E ∈ Eq(X) : $C\mathcal{U}$(E) = E}

and

Inv(G) = {E ∈ Eq(X) : x E y implies g(x) E g(y) for every g ∈ G}.

Outline

1. Introduction: From Object Certification to Object Selection

1.1 The completed stable-quotient foundation

1.2 What the previous theory does not yet determine

1.3 Supplied identity versus discovered identity

1.4 Object formation as constrained optimization

1.5 Compression, distortion, and lawful substitution

1.6 Why a variational principle is required

1.7 Principal contributions

1.8 Main results

2. The Search Space of Identity Policies

2.1 State spaces and candidate equivalence relations

2.2 The complete lattice Eq(X)

2.3 Fine and coarse identity policies

2.4 Quotient object spaces

2.5 Policy refinement as increased resolution

2.6 Policy coarsening as compression

2.7 Feasible and infeasible object systems

2.8 Optimization over partitions and quotients

3. Task-Relevant Distortion

3.1 Why erased differences must be measured

3.2 Observables and target quantities

3.3 Predictive distortion

3.4 Dynamical distortion

3.5 Causal and intervention distortion

3.6 Explanatory distortion

3.7 Worst-case and expected distortion

3.8 Zero-distortion quotients

4. Complexity and Compression of Object Systems

4.1 Quotient cardinality as elementary complexity

4.2 Logarithmic object complexity

4.3 Description length of equivalence policies

4.4 Structural complexity beyond class count

4.5 Compression ratios

4.6 Representation-independent complexity

4.7 Degenerate maximal compression

4.8 Why complexity must be balanced by distortion

5. Closure Coherence and Descent Constraints

5.1 The globalization–restriction operator

5.2 Closure operators on distributed policies

5.3 Coherent identity systems as fixed points

5.4 Descent defects as constraint violations

5.5 Hard versus penalized descent constraints

5.6 Local object preservation

5.7 Distributed optimization over observer networks

5.8 Closure-stable feasible regions

6. Transformation Invariance as an Admissibility Constraint

6.1 Admissible transformation families

6.2 Invariant equivalence relations

6.3 Congruence-like identity policies

6.4 Representative-independent quotient dynamics

6.5 Exact invariance

6.6 Approximate invariance

6.7 Invariance defects

6.8 The lattice of dynamically admissible policies

7. The Variational Objecthood Principle

7.1 Statement of the master optimization problem

7.2 Constrained and penalized formulations

7.3 Lagrange multipliers for object formation

7.4 The objecthood functional

7.5 Least-complexity admissible quotients

7.6 Maximum-compression formulations

7.7 Equivalent Pareto formulations

7.8 Interpretation as an algebraic least-action principle

8. Existence, Extremizers, and Finite Objecthood Theorems

8.1 Existence on finite state spaces

8.2 Compactness of finite policy lattices

8.3 Existence of minimizers

8.4 Nonuniqueness of optimal policies

8.5 Finest and coarsest minimizers

8.6 Tie-breaking by refinement

8.7 Stability under perturbation of λ

8.8 A finite variational objecthood theorem

9. Phase Transitions in Conceptual Resolution

9.1 The compression-distortion frontier

9.2 Critical values of the tradeoff parameter

9.3 Sudden object fusion

9.4 Sudden object birth

9.5 Plateaus of stable ontology

9.6 Bifurcations among optimal policies

9.7 Hysteresis and policy dependence

9.8 Emergence as an optimization phase transition

10. Learning and Discovering Identity Policies

10.1 From fixed labels to learned equivalence

10.2 Search over partitions

10.3 Local merge and split operators

10.4 Gradient-free policy optimization

10.5 Relaxed differentiable approximations

10.6 Constraint-guided clustering

10.7 Active observation for identity discovery

10.8 Artificial intelligence as self-revision of object space

11. Examples and Model Constructions

11.1 Finite classification with protected distinctions

11.2 Dynamical lumping of state-transition systems

11.3 Multisensor object formation

11.4 Scientific taxonomy under compression constraints

11.5 Image recognition under transformation invariance

11.6 Distributed identity with descent penalties

11.7 Emergent macrostates in physical systems

11.8 Computational search for optimal quotients

12. Variational Objecthood Theorem and Research Program

12.1 Statement of the combined theorem

12.2 Objecthood as constrained quotient optimization

12.3 Closure fixed points and invariant policies

12.4 Compression-distortion optimality

12.5 Existence and multiplicity of optimal ontologies

12.6 Implications for mathematics, science, and AI

12.7 Limits of scalar objecthood functionals

12.8 Next steps: graded identity, stochastic dynamics, and autonomous discovery

13. References

1. Introduction: From Object Certification to Object Selection

The preceding mathematics of object formation established a general sequence through which states become effective objects. A state space supplies available variation. An observation policy determines which states count as the same. The resulting equivalence classes form a quotient object space. Admissible transformations certify whether that quotient supports representative-independent dynamics. Refinement and coarsening organize changes in conceptual resolution. Distributed observation policies can be assembled through equivalence closure, while faithful descent determines whether global identity preserves locally maintained distinctions.

This foundation answers a major question:

When does a proposed identity policy produce lawful objects?

It does not yet answer the more difficult question:

Which identity policy should be proposed?

For a fixed state space X, there are generally many possible equivalence relations. Each relation produces a different quotient ontology. The equality relation treats every state as a separate object. The universal relation treats all states as one object. Between those extremes lie many possible classifications, abstractions, macrostates, representations, and conceptual resolutions.

Several of these policies may be mathematically coherent.

Several may be preserved by the same admissible transformations.

Several may descend faithfully across the same observer network.

The previous theory can certify each of them. It cannot by itself select one.

This absence of selection is not a minor technical gap. It marks the boundary between object certification and object discovery.

A scientist does not ask only whether a classification is internally consistent. The scientist also asks whether it is predictive, explanatory, economical, stable, and responsive to intervention. An artificial intelligence system does not become generally intelligent merely by preserving an externally supplied ontology. It must be capable of revising the categories through which its world is represented. A physical theory does not simply inherit macrostates. It determines which microscopic differences may be discarded without altering effective dynamics. A biological classification must balance diversity against common structure. A mathematical abstraction must suppress irrelevant detail while preserving the operations, relations, and invariants that matter.

The selection of an identity policy is therefore governed by a tradeoff.

A fine policy preserves detail but produces many objects.

A coarse policy compresses the state space but risks erasing distinctions required for prediction, explanation, intervention, or lawful transformation.

The preferred policy should retain no more detail than necessary, but it should not discard any distinction whose loss damages the intended task.

This is the central variational idea.

Object formation can be treated as optimization over the complete lattice of equivalence relations on a state space.

The candidate policies are the possible partitions of the state space.

The cost of a policy contains at least two components.

One component measures distortion: what task-relevant structure is lost when equivalent states are replaced by one quotient object?

The other measures complexity: how elaborate is the resulting object system?

A maximally fine policy produces minimal distortion but maximal object complexity.

A maximally coarse policy produces minimal object complexity but may create catastrophic distortion.

The useful object system normally lies between those extremes.

The central mathematical problem becomes:

Find the least complex identity policy that preserves the structure required for the intended prediction, explanation, intervention, or transformation.

This problem resembles rate-distortion theory, minimum description length, sufficient statistics, state aggregation, model reduction, clustering, abstraction, and coarse-graining. Yet it differs from each in scope. The object being optimized is not merely a code, representation, or statistical model. It is the policy of identity itself.

The optimization therefore ranges over possible effective ontologies.

A solution does not merely compress data. It determines which states become one object.

This paper develops that variational theory.

1.1 The Completed Stable-Quotient Foundation

Let X be a state space.

An identity policy is represented by an equivalence relation E on X.

The quotient

X/E

is the object space induced by E.

Each quotient class [x]E contains the states regarded as interchangeable under the selected policy.

The policy is stable under a transformation g: X → X when

x E y implies g(x) E g(y).

Under this condition, g induces a well-defined transformation on the quotient:

$\bar{g}([x]E) = [g(x)]E$.

Equivalent representatives therefore have equivalent images. The transformation of a quotient object does not depend on which underlying state is chosen to represent it.

If a family G of admissible transformations preserves E, the quotient supports representative-independent action by G. States within one equivalence class may differ internally, but those differences never create ambiguity at the quotient level.

This is the stable-quotient criterion.

The complete lattice Eq(X) organizes all identity policies on X.

For E, F ∈ Eq(X),

$E \subseteq F$

means that E is finer than F.

The finer policy preserves more distinctions.

The coarser policy identifies more states.

The meet of a family of policies is their intersection.

The join is the equivalence relation generated by their union.

The equality relation is the finest policy.

The universal relation is the coarsest.

Refinement splits quotient objects.

Coarsening merges them.

This lattice supplies the natural search space for variational objecthood.

The distributed theory adds a second requirement.

Suppose X is covered by observational domains Ui with local identity policies Ei. The local policies generate a global relation

Eglobal = EqCl(⋃i Ei).

Faithful descent occurs when

Eglobal restricted to Ui equals Ei

for every i.

Equivalently, the distributed family is unchanged by the operation that globalizes all local identities and then restricts the result back to each observer.

Write $C\mathcal{U}$ for this globalization–restriction operator. Then faithful descent is the fixed-point condition

$C\mathcal{U}$(E) = E.

The notation E here may represent an indexed family of local policies when the distributed setting is intended.

The completed foundational picture therefore contains three levels.

The first is quotient formation.

A policy E turns detailed states into effective objects.

The second is transformation stability.

Admissible transformations preserve E and descend to X/E.

The third is descent coherence.

Local policies combine into a global identity system without producing forbidden local fusion.

These levels supply necessary tests for objecthood.

They do not yet rank competing policies that pass those tests.

Suppose E and F are both invariant under G.

Suppose both are closure-coherent.

Suppose both produce lawful quotient dynamics.

If E has one thousand classes and F has ten, which policy is preferable?

The answer depends on what the quotient must preserve.

If both retain everything required by the task, F is normally preferable because it achieves greater compression.

If F erases a distinction required for prediction or intervention, E may be preferable despite its greater complexity.

The stable-quotient foundation therefore defines the admissible region of the problem.

Variational objecthood must define the objective within that region.

1.2 What the Previous Theory Does Not Yet Determine

The previous theory treats the observation policy as supplied.

A mathematician, scientist, observer, instrument, or algorithm provides an equivalence relation E. The theory then determines whether E produces coherent, stable, and globally realizable objects.

This leaves several questions unanswered.

Why was E selected rather than a finer policy?

Why was E selected rather than a coarser policy?

Which distinctions are essential?

Which distinctions are redundant?

How much compression can be achieved before predictive or explanatory damage appears?

When several admissible policies exist, is one preferable by a general mathematical principle?

Can an artificial system discover a useful quotient without receiving labels that already encode the intended objects?

Can object emergence be represented as a transition in the optimal policy rather than only as a verbal change of description?

These questions require a method for comparing identity policies.

The refinement order alone is insufficient.

A finer policy is not always better.

It preserves more information, but it also creates more objects, increases descriptive burden, and may conceal unifying structure.

A coarser policy is not always better.

It compresses more aggressively, but it may erase variables required for lawful behavior.

Transformation invariance is also insufficient.

Several policies may be invariant under the same dynamics.

The equality relation is invariant under every deterministic transformation. The universal relation is also invariant under every deterministic transformation. Both are stable in the formal sense, but neither is necessarily useful.

Faithful descent is insufficient for the same reason.

A distributed policy may be perfectly coherent while preserving far more detail than necessary.

Another coherent policy may collapse nearly everything.

Closure fixed points identify self-consistent ontologies. They do not identify informative or economical ones.

The missing ingredient is value relative to purpose.

A policy must be evaluated by what it preserves and what it costs.

Its cost may include the number of quotient objects, the complexity of describing the policy, the memory required to store the quotient, the computational burden of reasoning over its objects, or the structural complexity of its induced dynamics.

Its value may include predictive accuracy, causal fidelity, explanatory power, intervention reliability, symmetry preservation, dynamical autonomy, or successful transfer across representations.

The balance between preservation and economy creates an optimization problem.

Without that balance, the mathematics certifies objects but does not discover them.

1.3 Supplied Identity Versus Discovered Identity

A supplied identity policy enters the theory from outside.

Someone decides that two images represent the same object.

A physicist selects macroscopic variables.

A biologist defines a species criterion.

A database engineer specifies a record-matching rule.

A mathematician chooses equivalence up to isomorphism, homeomorphism, or homotopy.

The policy may be well motivated, but its origin remains external to the formalism.

A discovered identity policy is different.

The system begins with states, transformations, observables, tasks, constraints, or explanatory requirements. It searches for an equivalence relation that organizes those states into the most effective quotient objects.

The identity policy is then an output of the mathematics.

This shift is fundamental.

Under supplied identity, the theory asks whether a proposed object system is lawful.

Under discovered identity, the theory asks which lawful object system best satisfies a preservation objective.

The distinction parallels the difference between verifying a model and learning a model.

Verification begins with a candidate.

Learning selects among candidates.

The object-selection problem is more difficult than ordinary classification because the labels are not assumed in advance. The system must determine how many objects should exist, which states belong together, and which distinctions should survive.

Clustering partially addresses this problem, but ordinary clustering usually begins with a distance, similarity measure, number of clusters, or parametric family supplied externally. It does not necessarily test transformation invariance, descent coherence, causal preservation, or representative-independent dynamics.

Variational objecthood requires these features to be considered together.

The discovered policy should not merely group nearby states.

It should license substitution.

If x E y, the system asserts that x and y may be treated as the same object for the intended purpose.

That assertion must remain valid under the relevant observations and operations.

Discovered identity is therefore stronger than detected similarity.

Similarity may be local, graded, asymmetric, or nontransitive.

Identity is a global rule of admissible replacement.

The optimization problem must determine when similarity is sufficiently strong and structurally coherent to become equivalence.

This is also the point at which artificial intelligence can move beyond fixed ontologies.

A system with fixed labels learns mappings into an object space selected by humans.

A system capable of discovered identity can reorganize the object space itself.

It can split a category when hidden distinctions become predictive.

It can merge categories when their differences cease to matter.

It can revise its ontology in response to new transformations, goals, or evidence.

Such a system does not merely learn within a representation.

It learns the representation's units of objecthood.

1.4 Object Formation as Constrained Optimization

Let $Eq(X)$ be the complete lattice of candidate identity policies on X.

For each $E \in Eq(X)$, define a cost functional

$J_{\pi,\lambda}(E)$.

The symbol π represents the task, explanatory regime, or admissibility policy.

The parameter λ controls the tradeoff between preservation and simplicity.

A basic form is

$J_{\pi,\lambda}(E) = D_{\pi}(E) + \lambda K(E)$,

where $D_{\pi}(E)$ is the task-relevant distortion introduced by quotienting under E, and $K(E)$ is the complexity of the resulting object system.

The preferred policy is selected by minimizing this cost over an admissible family of equivalence relations.

Let A_{π} be the feasible set of policies satisfying the required structural constraints. Then

$E^{\star} \in \arg\min \{D_{\pi}(E) + \lambda K(E) : E \in A_{\pi}\}$.

The feasible set may require several conditions.

A transformation-invariance condition may require

x E y implies g(x) E g(y)

for every admissible g ∈ G.

A descent-coherence condition may require

C$\mathcal{U}$(E) = E.

A protected-distinction condition may prohibit specified pairs from becoming equivalent.

An empirical condition may require selected observables, measurements, or causal responses to be preserved.

Thus one may define

Aπ = {E ∈ Eq(X) : E is closure-coherent, transformation-invariant, and compatible with all protected distinctions}.

The optimization then takes place only among identity systems already certified as structurally lawful.

Several equivalent formulations are possible.

One may minimize complexity subject to a distortion bound:

minimize K(E)

subject to

Dπ(E) ≤ ε.

One may minimize distortion subject to a complexity budget:

minimize Dπ(E)

subject to

K(E) ≤ B.

One may maximize compression subject to exact preservation of selected observables.

One may maximize predictive value minus a complexity penalty.

One may search for Pareto-optimal policies for which no alternative improves both distortion and complexity.

These formulations answer different scientific questions.

The parameter λ has a direct conceptual interpretation.

When λ is small, distortion is expensive relative to complexity. The selected policy tends to remain fine.

When λ is large, complexity is expensive relative to distortion. The selected policy tends to become coarse.

As λ varies, the optimal equivalence relation may change discontinuously.

Several quotient classes may merge at a critical value.

A formerly unified class may split when distortion becomes too costly.

Conceptual resolution can therefore undergo phase transitions.

Emergence can be represented as a change in the optimal quotient rather than only as a change in descriptive language.

The variational formulation also allows policies to be compared when neither refines the other.

Two incomparable partitions may preserve different structures.

Their costs determine which is preferable for the selected purpose π.

Optimization therefore supplements the lattice order.

The lattice describes possible changes in identity.

The functional evaluates those changes.

1.5 Compression, Distortion, and Lawful Substitution

Every quotient performs compression.

Several states become one object.

If X is finite, a simple measure of quotient complexity is

$K(E) = \log |X/E|$.

The equality policy gives

$K(\Delta X) = \log |X|$.

The universal policy gives

$K(\nabla X) = 0$.

This measure represents the idealized information required to specify one quotient object under a uniform code.

More sophisticated complexity measures may include the description length of the equivalence relation, the structural complexity of the quotient dynamics, the number of parameters required to represent the classes, or the computational cost of deciding whether two states are equivalent.

Compression alone favors the universal relation.

If all states become one object, quotient complexity is minimized.

That policy is normally useless because it erases every distinction.

Distortion measures the consequence of that erasure.

Suppose $h: X \to Y$ is a task-relevant observable.

If $x \, E \, y$ but $h(x) \neq h(y)$, then the quotient cannot represent h exactly.

A worst-case distortion may be defined as the largest distance between observable values assigned to states in one class:

$D\pi(E) = \sup \{dY(h(x), h(y)) : x \, E \, y\}$.

An expected distortion may average over states or equivalent pairs according to a probability distribution.

For prediction, h may represent future outcomes.

For causal analysis, h may represent intervention responses.

For scientific explanation, h may represent quantities the theory must preserve.

For dynamics, distortion may compare trajectories generated by equivalent initial states.

A zero-distortion policy satisfies

$x \, E \, y$ implies $h(x) = h(y)$.

Then h factors through the quotient:

$h = \bar{h} \circ qE$.

The quotient contains all information required to determine h.

This factorization is the mathematical expression of sufficient abstraction.

The classes erase many differences, but none of those differences affect the selected observable.

Lawful substitution combines this principle with transformation invariance.

If x E y, one may replace x by y without changing the task-relevant result or the quotient-level dynamics.

The equivalence class becomes a valid unit of reasoning.

The preferred quotient should therefore maximize the scope of admissible substitution while limiting distortion.

The strongest possible compression consistent with exact task preservation is the coarsest equivalence relation contained in the kernel of every required observable and preserved by every required transformation.

In simple cases, that relation can be constructed directly.

In more complex cases, several observables, transformations, local policies, and approximate constraints interact. Optimization becomes necessary.

The objecthood functional measures the price of declaring differences irrelevant.

Compression rewards broader substitution.

Distortion penalizes substitutions that alter what matters.

A stable object is an equivalence class whose internal differences remain irrelevant under the chosen task and transformation policy.

1.6 Why a Variational Principle Is Required

A variational principle selects structure by minimizing or maximizing a functional.

Classical mechanics selects trajectories through stationary action.

Thermodynamics selects equilibrium states through extremization of thermodynamic potentials under constraints.

Information theory selects codes and channels through entropy and distortion objectives.

Statistics selects models through likelihood, risk, posterior probability, or description length.

A mathematics of object formation requires an analogous selection principle.

Without one, the theory remains permissive.

It describes many possible ontologies but does not explain why one becomes effective.

A variational principle is required for five principal reasons.

First, admissibility alone leaves extreme solutions.

Equality and universal equivalence are frequently both admissible.

Optimization explains why useful object systems normally occupy intermediate resolutions.

Second, objecthood is purpose-relative.

The same state space may support different optimal quotients for prediction, control, explanation, communication, or intervention.

The objective functional makes that purpose explicit.

Third, emergence involves selection among scales.

A macro-object becomes meaningful when its quotient provides substantial compression while retaining sufficiently autonomous dynamics.

This is naturally expressed as an optimum.

Fourth, artificial intelligence requires policy discovery.

An AI system capable of revising its ontology needs a criterion for evaluating proposed mergers and splits.

Fifth, scientific concepts are judged by tradeoffs.

A successful concept unifies many cases without erasing distinctions required by evidence.

This is precisely a compression-distortion balance.

A variational principle also explains why one elegant governing equation did not appear in the earlier papers.

Those papers developed the search space and the constraints.

They identified the candidate policies, the closure operator, the invariance conditions, the quotient structures, and the descent criteria.

A governing equation becomes possible only after an objective functional is introduced.

The variational equation gives direction to the framework.

Instead of saying only which policies are allowed, it identifies which allowed policies are preferred.

The resulting principle can be stated informally:

Among all identity policies that preserve what matters, choose the one that produces the least complex effective object system.

In approximate form:

Choose the identity policy that achieves the best balance between compression and task-relevant distortion.

This is the proposed least-complexity principle of objecthood.

1.7 Principal Contributions

The paper makes eight principal contributions.

First, it identifies Eq(X), the complete lattice of equivalence relations on a state space, as the natural search space for object selection.

Second, it introduces task-relevant distortion functions on identity policies. These functions measure the consequences of erasing distinctions through quotienting.

Third, it defines complexity functionals for quotient object systems, including cardinal, logarithmic, descriptive, structural, and computational measures.

Fourth, it formulates object formation as constrained optimization over Eq(X).

The fundamental problem is to minimize

$D\pi(E) + \lambda K(E)$

over policies satisfying stability, descent, and admissibility constraints.

Fifth, it defines coherent distributed identity systems as fixed points of the globalization–restriction operator and incorporates descent defects as either hard constraints or penalty terms.

Sixth, it treats transformation invariance as part of the feasible region and introduces invariance-defect terms for applications in which exact preservation is too strong.

Seventh, it establishes existence of optimal policies on finite state spaces and studies nonuniqueness, finest and coarsest minimizers, tie-breaking rules, and changes in optimal resolution as λ varies.

Eighth, it interprets changes in the optimizing policy as phase transitions in objecthood.

Sudden class mergers represent object fusion.

Sudden refinements represent object birth.

Stable parameter intervals represent persistent ontological regimes.

The paper also establishes a research program for autonomous concept discovery.

A computational system can search the partition lattice, evaluate candidate quotients, detect descent and stability violations, and revise its identity policy through local merge and split operations.

This creates a path from object certification to object generation.

1.8 Main Results

The central structure can now be summarized.

Let X be a finite or otherwise suitably controlled state space.

Let Eq(X) be its lattice of identity policies.

Let π specify the task-relevant observables, predictions, interventions, or explanatory quantities.

Let G be the family of admissible transformations.

Let $C\mathcal{U}$ be the globalization–restriction closure operator associated with a distributed observational system.

Define the family of transformation-invariant policies by

Inv(G) = {E ∈ Eq(X) : x E y implies g(x) E g(y) for every g ∈ G}.

Define the family of closure-coherent policies by

Fix($C\mathcal{U}$) = {E ∈ Eq(X) : $C\mathcal{U}$(E) = E}.

Additional protected distinctions or empirical constraints may further restrict the feasible family.

Define the objecthood cost

Jπ,λ(E) = Dπ(E) + λK(E).

An optimal object policy is any equivalence relation satisfying

E★ ∈ arg min {Dπ(E) + λK(E) : E ∈ Fix($C\mathcal{U}$) ∩ Inv(G)}.

When protected distinctions or other admissibility conditions are present, they are added to the set over which the minimum is taken.

This equation joins the three recent mathematical developments.

The fixed-point condition expresses coherent local-to-global identity.

Transformation invariance expresses stable quotient dynamics.

The variational objective selects the most economical policy compatible with the task.

For finite X, the lattice Eq(X) is finite. If the feasible set is nonempty, at least one minimizer exists.

The minimizer need not be unique.

Several policies may achieve the same cost while producing different quotient ontologies.

The set of minimizers can itself be ordered by refinement.

A finest minimizer preserves the greatest number of distinctions among optimal policies.

A coarsest minimizer achieves the greatest compression among policies with the same optimal cost.

Additional tie-breaking principles may select between them.

If $D\pi$ is zero on a nonempty feasible family, the exact least-complexity problem becomes

$E\bigstar \in \arg\min \{K(E) : E \text{ is feasible and } D\pi(E) = 0\}$.

This selects a maximally compressed lawful quotient that preserves the task exactly.

As λ varies, the minimizing policy changes only at critical parameter values in the finite case.

Between critical values, one policy or one family of policies remains optimal.

At a critical value, two or more policies exchange optimality.

The quotient ontology may then change discontinuously.

This produces a piecewise-stable path of optimal resolutions:

$\lambda \mapsto E\bigstar\lambda$.

The path is not necessarily monotone without additional assumptions, although natural distortion and complexity functionals often favor coarsening as λ increases.

The central variational principle can therefore be stated as:

$E\bigstar \in \arg\min \{D\pi(E) + \lambda K(E) : E \in \mathrm{Fix}(C\mathcal{U}) \cap \mathrm{Inv}(G)\}$.

Objecthood is no longer represented merely by an equivalence class.

It is represented by an optimized equivalence policy.

The effective objects are the classes of the policy selected by the least-complexity preservation principle.

The conceptual sequence is now complete:

states become equivalent,

equivalence produces quotient objects,

stability certifies their transformations,

descent certifies their distributed coherence,

and optimization selects which certified object system should exist at the effective level.

This is the transition from object certification to object selection.

It is also the point at which the mathematics of concept formation acquires a genuine variational law.

2. The Search Space of Identity Policies

A variational theory requires a clearly defined space over which optimization takes place. In the present framework, the candidates are not trajectories, parameter vectors, probability distributions, or numerical states. They are identity policies.

An identity policy determines which states may be substituted for one another without leaving the intended conceptual resolution. Mathematically, such a policy is represented by an equivalence relation on a state space X.

The collection of all equivalence relations on X is denoted

Eq(X).

Every element $E \in Eq(X)$ produces a quotient object space

X/E.

Thus the search over identity policies is simultaneously a search over possible object systems.

Different policies generate different ontologies from the same underlying states.

One policy may preserve every distinction.

Another may merge states according to common predictions.

Another may identify states related by symmetry.

Another may group them by intervention response.

Another may collapse nearly all variation.

The state space alone does not determine which of these object systems is preferred. It supplies the material from which object systems may be formed. The policy determines which differences remain visible.

The search space Eq(X) has substantial internal structure. It is not merely an unorganized collection of partitions. It is a complete lattice under refinement. Every family of identity policies has a greatest common refinement and a least common coarsening. This lattice structure organizes object fusion, object splitting, conceptual resolution, and the possible paths through which an optimization algorithm can search.

The finest policy is equality. It retains every state as a separate object.

The coarsest policy is universal equivalence. It treats the entire state space as one object.

All nontrivial object systems lie between these extremes.

This section develops Eq(X) as the foundational search space of variational objecthood. It defines candidate relations, quotient spaces, refinement order, feasible regions, policy constraints, and the computational problem of optimization over partitions.

2.1 State Spaces and Candidate Equivalence Relations

Let X be a nonempty set of states.

A state may represent a physical configuration, image, data record, biological specimen, mathematical presentation, computational state, scientific model, observation, or any other element capable of being distinguished before quotienting.

An equivalence relation E on X is a binary relation satisfying three conditions.

Reflexivity:

$x \mathrel{E} x$

for every $x \in X$.

Symmetry:

$x \mathrel{E} y$ implies $y \mathrel{E} x$.

Transitivity:

x E y and y E z imply x E z.

The equivalence class of x is

[x]E = {y ∈ X : y E x}.

The classes of E partition X.

They are pairwise disjoint, nonempty, and their union is X.

Conversely, every partition of X determines a unique equivalence relation. Two states are equivalent exactly when they belong to the same partition block.

Thus equivalence relations and partitions are two presentations of the same identity policy.

The relational presentation emphasizes substitution.

The partition presentation emphasizes object classes.

The quotient presentation emphasizes the resulting object space.

All three are useful.

A candidate identity policy is any E ∈ Eq(X) admitted for evaluation by the variational theory.

At the broadest level, every equivalence relation is a candidate.

In practical applications, additional structure may restrict the candidates.

Some policies may violate known symmetries.

Some may fail transformation invariance.

Some may merge states with incompatible outcomes.

Some may contradict protected distinctions.

Some may fail local-to-global descent.

Some may be too complex to represent or compute.

The unrestricted lattice Eq(X) is therefore the ambient search space.

The admissible search space is a subset

Aπ ⊆ Eq(X)

determined by the task and constraints.

The role of the variational theory is to compare elements of Aπ.

The identity policy must not be confused with a similarity relation.

Similarity may be graded.

It may fail transitivity.

It may depend on scale.

It may be asymmetric.

An equivalence relation makes a stronger assertion.

It declares that the states in one class are interchangeable for the purposes encoded by the policy.

This distinction is essential.

Suppose images x and y are visually similar, while y and z are also visually similar, but x and z are visibly different. Similarity need not be transitive.

If all three pairs are converted into equivalence through closure, x and z become one object.

That conclusion is justified only if the intended identity policy licenses the entire chain.

The search over Eq(X) therefore occurs after the theory has decided that objecthood is represented by lawful substitution rather than by mere proximity.

Approximate or graded identity will require later extensions.

The present theory begins with exact equivalence because it provides an unambiguous quotient space and a precise optimization domain.

A policy E may be generated in several ways.

It may be supplied directly as a partition.

It may be generated by a family of identity declarations.

It may be the kernel of an observable

$h: X \to Y$,

so that

$x \, E \, y$ exactly when $h(x) = h(y)$.

It may be the orbit relation of a symmetry group acting on X.

It may be a behavioral equivalence defined by identical predictions or intervention responses.

It may be a bisimulation relation on a transition system.

It may be the closure of local observer policies.

It may also be found by minimizing an objecthood functional.

The variational framework is neutral concerning the source of the candidates. It requires only that each candidate determine a lawful equivalence relation.

2.2 The Complete Lattice Eq(X)

The set Eq(X) is ordered by inclusion.

For $E, F \in Eq(X)$, define

$E \preceq F$

when

$E \subseteq F$.

Under this convention, E is finer than F.

Every pair identified by E is also identified by F, but F may identify additional pairs.

Thus moving upward in the order means coarsening the policy.

Moving downward means refining it.

The same order can be described in terms of quotient classes.

$E \preceq F$ exactly when every E-class is contained in one F-class.

Therefore the partition induced by E refines the partition induced by F.

The relation $\preceq$ is a partial order.

It is reflexive because $E \subseteq E$.

It is antisymmetric because $E \subseteq F$ and $F \subseteq E$ imply $E = F$.

It is transitive because $E \subseteq F$ and $F \subseteq H$ imply $E \subseteq H$.

The finest policy is the equality relation

$\Delta X = \{(x, x) : x \in X\}$.

Every state is its own class.

For every $E \in Eq(X)$,

$\Delta X \preceq E$.

The coarsest policy is the universal relation

$\nabla X = X \times X$.

All states belong to one class.

For every $E \in Eq(X)$,

$E \preceq \nabla X$.

The meet of a family $\{E\alpha\}\alpha \in A$ is their intersection:

$\bigwedge \alpha\, E\alpha = \bigcap \alpha\, E\alpha$.

The intersection of equivalence relations is again an equivalence relation.

It contains exactly the identity judgments accepted by every member of the family.

The meet is the greatest policy finer than all $E\alpha$.

Its classes preserve every distinction preserved by at least one policy in the family.

The join of the family is

$\bigvee \alpha\, E\alpha = EqCl(\bigcup \alpha\, E\alpha)$.

The union of equivalence relations need not be transitive and may therefore fail to be an equivalence relation. Equivalence closure is required.

The join contains every identity accepted by any $E\alpha$, together with every identity forced by transitivity.

It is the least policy coarser than all $E\alpha$.

Thus $Eq(X)$ is a complete lattice.

Every family of policies, finite or infinite, has a meet and a join.

This completeness is important for optimization.

It means that arbitrary collections of identity policies can be combined through common refinement or generated coarsening.

It also supports fixed-point methods, closure operators, monotone dynamics, and order-theoretic search.

The lattice may be extremely large.

If X has n finite elements, the number of equivalence relations is the nth Bell number.

For small n, the growth is already rapid.

A state space with 1 element has 1 partition.

A state space with 2 elements has 2.

A state space with 3 elements has 5.

A state space with 4 elements has 15.

A state space with 5 elements has 52.

A state space with 10 elements has 115,975 possible partitions.

The search space therefore becomes combinatorially enormous long before the state space itself becomes large.

This fact has two consequences.

First, the mathematical formulation is exact but exhaustive computation may be impractical.

Second, the lattice structure must be exploited.

Search procedures should move through local refinements and coarsenings rather than enumerate every policy.

The lattice also clarifies what it means for an object ontology to change.

A downward move splits one or more classes.

An upward move merges one or more classes.

A move between incomparable policies reorganizes the classes without being purely a refinement or coarsening.

The variational landscape is therefore not a linear sequence of resolutions. It is a branching order of possible object systems.

2.3 Fine and Coarse Identity Policies

A fine identity policy preserves many distinctions.

A coarse policy suppresses many distinctions.

These terms are relative.

Given E and F, one can say that E is finer than F only when $E \subseteq F$.

If neither relation contains the other, the policies are incomparable.

The equality relation is maximally fine.

Its quotient is naturally identified with X:

$X/\Delta X \cong X$.

No compression occurs.

Every state remains a separate object.

This policy has several advantages.

It preserves every observable.

It is stable under every deterministic transformation.

It never confuses distinct states.

It also provides no abstraction.

The quotient is as complex as the original state space.

The universal relation is maximally coarse.

Its quotient contains one class:

$X/\nabla X = \{X\}$.

Maximum compression occurs.

Every state becomes one object.

This policy is also stable under every transformation.

It is useless for any task requiring distinctions among states.

These two extreme policies show that stability alone does not select a meaningful ontology.

Both are always invariant.

The variational functional must prefer intermediate policies when the task rewards compression but penalizes excessive information loss.

A fine policy may be appropriate when small differences have important consequences.

Medical diagnosis may require separating states that appear nearly identical but respond differently to treatment.

Control systems may require distinctions among states leading to different future trajectories.

A proof system may need to preserve syntactic distinctions when substitutions are not valid.

A coarse policy may be appropriate when many differences do not affect the intended task.

Microscopic particle arrangements may be irrelevant to a thermodynamic description.

Different image poses may represent the same physical object.

Several program states may be equivalent when they produce the same externally visible behavior.

Different mathematical presentations may define isomorphic structures.

The correct resolution is therefore policy-dependent.

Fine and coarse should not be interpreted as good and bad.

They describe the amount of distinction retained.

A policy is useful when its resolution matches the distinctions required by the task.

There are two common mistakes.

The first is over-refinement.

The system preserves distinctions that have no predictive, explanatory, causal, or operational value.

This creates unnecessary object proliferation.

The object system becomes difficult to learn, store, communicate, and reason over.

The second is over-coarsening.

The system merges states whose differences matter.

This creates distortion, ambiguity, unstable quotient dynamics, or failed interventions.

Variational objecthood balances these risks.

Refinement tends to reduce distortion and increase complexity.

Coarsening tends to reduce complexity and increase distortion.

These tendencies need not hold strictly for every possible functional, but they motivate the central tradeoff.

Let $E \preceq F$.

If K measures quotient class count or a monotone version of representational complexity, then one expects

$K(F) \leq K(E)$.

The coarser policy F has no more classes than E.

If $D\pi$ penalizes task-relevant variation within classes, then one commonly expects

$D\pi(E) \leq D\pi(F)$.

The finer policy E cannot introduce more within-class variation than F because its classes are subsets of F-classes.

Under such monotonicity assumptions, the refinement order aligns directly with the compression-distortion tradeoff.

The optimization then seeks a point at which further coarsening produces too much distortion and further refinement produces unnecessary complexity.

2.4 Quotient Object Spaces

Every identity policy E produces a quotient map

$qE: X \rightarrow X/E$

defined by

$qE(x) = [x]E$.

The quotient map sends each state to its effective object.

It is surjective.

Every quotient class contains at least one state.

The quotient map is generally not injective. Its failure of injectivity is exactly the compression created by E.

For $x, y \in X$,

$qE(x) = qE(y)$

if and only if

$x \, E \, y$.

Thus E is the kernel relation of its quotient map.

The quotient object space X/E is the effective ontology determined by E.

Its elements are not additional states placed beside X.

They are classes of states under a licensed substitution policy.

The quotient may be much smaller than X.

If X is finite,

$|X/E|$

is the number of effective objects.

If $E \preceq F$, there is a canonical surjection

$p_{E,F}: X/E \to X/F$

given by

$p_{E,F}([x]_E) = [x]_F$.

The map is well defined because $E \subseteq F$.

An E-class lies inside one F-class.

This canonical map expresses further abstraction.

The path

$X \to X/E \to X/F$

is equivalent to quotienting X directly by F.

Thus every coarsening can be represented as a quotient of an existing quotient.

Object fusion is compositional.

Refinement behaves differently.

If $E \preceq F$, there is generally no canonical map from X/F to X/E. One F-object contains several E-objects, and selecting one would require additional information.

Coarsening forgets distinctions.

Refinement must restore distinctions that the coarser quotient no longer contains.

This asymmetry is central to learning and concept revision.

Merging objects is easy to define.

Splitting an object requires access to underlying states, stored provenance, or new observations.

A computational system that discards all internal information after quotienting may be unable to refine its ontology later.

A variational object system should therefore distinguish the effective quotient from the retained evidence supporting it.

The quotient may be used for reasoning while the underlying states remain available for revision.

The quotient also supports factorization.

Let $h: X \to Y$ be a function.

The function h factors through X/E exactly when it is constant on E-classes:

x E y implies h(x) = h(y).

When this condition holds, there exists a unique function

$\bar{h}: X/E \to Y$

such that

$h = \bar{h} \circ qE$.

This universal property expresses preservation.

The quotient retains exactly enough information for h when h is constant on its classes.

If several observables h1, …, hm must be preserved, the policy must refine the intersection of their kernel relations.

A quotient capable of representing all required observables must not merge states that any required observable distinguishes.

This produces a direct method for constructing exact feasible policies.

Approximate preservation replaces exact constancy with bounded distortion.

The quotient then represents h only approximately.

The variational framework quantifies the error.

2.5 Policy Refinement as Increased Resolution

Let $E, F \in Eq(X)$ with $E \preceq F$.

The passage from F to E is refinement.

Every F-class is divided into one or more E-classes.

The object system gains resolution.

Refinement may be driven by new observations.

Two states previously treated as one object may be found to have different future behavior.

A medical category may split into treatment-responsive and treatment-resistant subtypes.

A biological species may divide into distinct lineages.

A scientific phase may split when an order parameter becomes measurable.

An AI system may separate one concept into several because the previous category produces prediction errors.

Mathematically, refinement removes identity pairs.

Because an equivalence relation must remain reflexive, symmetric, and transitive, pairs cannot always be removed independently.

A class must be repartitioned into smaller classes.

A refinement can therefore be represented by a partition of one or more existing quotient objects.

Refinement tends to improve fidelity.

If an observable varies within an F-class, splitting that class may reduce distortion.

If a quotient transformation is ambiguous because representatives have different outcomes, refinement can restore stability.

If a global class contains two locally distinguished objects, refinement of one or more generating policies can repair descent.

Refinement also increases complexity.

The number of quotient objects cannot decrease.

The descriptive burden may rise.

The quotient dynamics may become more detailed.

Data requirements may increase because the system must decide among more classes.

Thus refinement is not free.

The variational objective should permit refinement only when the reduction in distortion justifies the increase in complexity.

A local refinement operation can be defined by selecting one class $C \in X/F$ and replacing it with a partition

$C = C1 \cup \cdots \cup Ck.$

All other classes remain unchanged.

The resulting relation E is immediately finer than F.

Such local split operations provide a natural neighborhood structure for search algorithms.

One may evaluate the change in objective

$\Delta J = J(E) - J(F).$

If $\Delta J < 0$, the refinement improves the variational cost.

If $\Delta J > 0$, the added resolution is not justified under the current weighting.

Multiple local refinements may interact.

A split that appears unhelpful alone may enable another split or remove a large stability defect.

The objective landscape may therefore contain local minima.

Search procedures may require coordinated moves, annealing, branch-and-bound methods, or relaxations.

Refinement can also change feasibility.

A coarse policy may violate a protected distinction.

Splitting the offending class can restore admissibility.

A coarse policy may fail transformation invariance in a more subtle way: equivalent states may evolve into different quotient classes. Refinement can separate those states before transformation and make the quotient dynamics well defined.

Thus increased resolution is one of the principal repair operations of the theory.

2.6 Policy Coarsening as Compression

The passage from E to F with $E \preceq F$ is coarsening.

Several E-classes are merged into one F-class.

The object system becomes simpler.

Coarsening removes distinctions.

It increases the range of lawful substitution claimed by the policy.

If x and y belonged to different E-classes but one F-class, the coarser theory treats them as interchangeable.

Coarsening may be justified when the distinction between those objects does not affect any task-relevant quantity.

Two detailed physical configurations may have identical macroscopic behavior.

Two mathematical constructions may differ in presentation but be isomorphic.

Two image states may differ by lighting or viewpoint while representing the same object.

Two internal computational states may have the same future observable behavior.

When the merger is lawful, the quotient becomes smaller without meaningful loss.

This is abstraction.

A local coarsening move selects several existing classes and merges them.

If E has classes A and B, a simple merge creates the least equivalence relation F containing E and all pairs in $A \times B$.

The classes A and B become $A \cup B$.

All other classes remain unchanged.

For finite partitions, this is a basic search operation.

The change in objective is again

$\Delta J = J(F) - J(E)$.

The merge reduces complexity.

It may increase distortion.

It may also create new descent or stability violations.

Because equivalence is transitive, one merge can have consequences beyond the selected classes when the policy is embedded in a distributed or typed system.

A new bridge may connect previously separate global components.

Coarsening is therefore potentially amplifying.

A seemingly modest identity declaration may trigger large fusion after closure.

The algorithm must recompute or update all relevant constraints.

Coarsening is often the mechanism by which emergent objects appear.

A collection of lower-level states becomes one macro-object when their internal differences cease to matter relative to the selected task.

The macro-object is not an extra entity added to the state space.

It is the quotient class created by a successful coarsening.

This does not make emergence arbitrary.

The merger must satisfy the preservation constraints and improve the objective.

An emergent object is variationally justified when the compression achieved by its class exceeds the distortion introduced by treating its members as one.

The concept of sufficient coarsening is useful.

A policy is insufficiently coarse when additional lawful mergers remain possible without increasing task distortion beyond the allowed level.

A policy is excessively coarse when one or more mergers erase required distinctions.

The optimal policy lies at the boundary between these regimes.

2.7 Feasible and Infeasible Object Systems

Not every equivalence relation is admissible for every problem.

Let A_π denote the feasible set of identity policies under task policy π.

A policy E belongs to A_π only if it satisfies all hard constraints.

The constraints may include exact observable preservation.

For a family of required observables H,

$x \mathrel{E} y$ must imply $h(x) = h(y)$

for every $h \in H$.

Equivalently,

$E \subseteq \bigcap_{h \in H} \mathrm{Ker}(h)$.

A policy may also be required to preserve transformation dynamics.

For every $g \in G$,

$x \, E \, y$ must imply $g(x) \, E \, g(y)$.

This places E in Inv(G).

A distributed policy may be required to satisfy closure coherence:

$C_{\mathcal{U}}(E) = E$.

Protected distinctions may impose

$x \text{ not } E \, y$

for selected pairs (x, y).

These may represent experimentally certified differences, simultaneous object separation, causal incompatibility, logical disequality, or safety constraints.

A policy may also face cardinality or complexity bounds:

$|X/E| \leq B$

or

$K(E) \leq B$.

Alternatively, minimum-resolution constraints may require at least a specified number of classes.

The feasible region is the intersection of all such requirements.

For example,

$A_\pi = \mathrm{Fix}(C_{\mathcal{U}}) \cap \mathrm{Inv}(G) \cap \mathrm{Sep}(P) \cap \mathrm{Pres}(H)$,

where Sep(P) is the family preserving all protected distinctions P, and Pres(H) is the family preserving all observables H.

A policy outside A_π is infeasible.

It may be infeasible for several reasons.

It may merge states with different required outputs.

It may fail representative-independent dynamics.

It may create a gluing defect.

It may violate a cannot-link constraint.

It may exceed a resource budget.

It may fail to satisfy structural requirements imposed by the application.

The feasible set may be empty.

This is important.

An empty feasible set means that the constraints themselves cannot be satisfied simultaneously by any equivalence relation.

The response should not be to select the least bad policy without acknowledging inconsistency.

The system must determine which constraints should be relaxed, revised, contextualized, or reweighted.

Feasibility analysis therefore precedes optimization.

In a hard-constrained formulation, no infeasible policy may be selected regardless of its compression benefit.

In a penalized formulation, violations contribute additional cost.

One may define

$J(E) = D\pi(E) + \lambda K(E) + \mu S(E) + \nu L(E)$,

where $S(E)$ measures stability violation and $L(E)$ measures descent defect.

Large values of μ and ν approximate hard constraints.

Finite values permit tradeoffs.

The choice between hard and soft constraints is substantive.

A protected legal identity distinction may be nonnegotiable.

A noisy approximate invariance condition may be better represented by a penalty.

A scientific theory may tolerate small predictive error but not logical contradiction.

The formalism must therefore identify which requirements are absolute and which are graded.

Feasibility is also policy-relative.

A relation may be infeasible for prediction but feasible for compression.

It may be appropriate for one intervention and not another.

It may be coherent under one observer cover and obstructed under another.

There is no universal feasible set independent of task and observational context.

The notation A_π emphasizes this dependence.

2.8 Optimization over Partitions and Quotients

The variational objecthood problem is an optimization over $Eq(X)$ or over a feasible subset A_π.

The basic problem is

$$E^\star \in \arg\min \{D_\pi(E) + \lambda K(E) : E \in A_\pi\}.$$

When X is finite, $Eq(X)$ is finite.

Therefore a minimizer exists whenever A_π is nonempty.

This existence result is elementary but important. The finite theory always has at least one optimal identity policy under a real-valued objective.

The computational problem may nevertheless be difficult because the number of partitions grows rapidly.

Several search strategies are available.

Exhaustive search evaluates every feasible partition.

This is exact but practical only for small state spaces.

Agglomerative search begins with the equality policy and repeatedly merges classes.

At each step, it selects the merger producing the largest improvement or smallest increase in objective.

This creates a monotone path toward coarser policies.

Divisive search begins with the universal policy and repeatedly splits classes.

It is appropriate when coarse policies are easy to initialize but require repair.

Bidirectional search permits both merges and splits.

It can escape some errors introduced by one-way greedy procedures.

Lattice branch-and-bound methods use lower and upper bounds on the objective to discard large regions of $Eq(X)$.

Dynamic programming may be possible when the state space or objective has special decomposable structure.

Constraint satisfaction methods can encode required and forbidden equivalences.

Integer programming can represent class assignments and pairwise identity variables.

Graph methods can search over connected components generated by selected edges.

Probabilistic methods can sample partitions according to posterior or energy-like weights.

Differentiable relaxations can replace hard equivalence assignments with soft matrices before projecting the result back to a partition.

Each method searches an approximation or portion of the same underlying lattice.

The search should respect the distinction between a partition representation and an equivalence relation representation.

Partition representations store classes directly.

Pairwise representations store binary variables

$e(x, y) \in \{0, 1\}$.

The pairwise variables must satisfy reflexivity, symmetry, and transitivity.

Transitivity creates a large family of constraints:

$e(x, y) + e(y, z) - 1 \leq e(x, z)$.

These constraints make optimization nonlocal.

A local decision to identify x with y can force many additional identifications.

This is the computational expression of closure.

For large problems, one may search over generators rather than complete relations.

A candidate graph of identity edges is selected.

Its connected components define the equivalence classes.

The objective is then evaluated after closure.

This allows sparse representations, but it also creates redundancy because many different edge sets generate the same equivalence relation.

Optimization should ultimately compare the generated policies, not merely their presentations.

The quotient representation may reduce the problem further.

If the current policy has classes C1, …, Ck, local coarsening can be searched over merges among these quotient objects rather than over individual states.

Refinement requires returning to internal class structure.

Thus an efficient algorithm may alternate between state-level analysis and quotient-level moves.

The optimization problem also has a geometric interpretation.

Each equivalence relation is a point in a discrete policy space.

Refinement edges connect policies differing by one elementary split or merge.

The objective J assigns a height to every point.

Optimal ontologies are minima in this landscape.

Critical transitions occur when changes in λ alter which policy has the lowest height.

Local minima represent policies resistant to small changes but not necessarily globally optimal.

Flat regions represent families of equally valued ontologies.

The search landscape may contain many symmetries.

Relabeling quotient classes does not change the policy.

Equivalent representations should not be counted as distinct candidates.

The partition lattice automatically removes this label redundancy.

The final output of optimization should include more than E★.

A rigorous system should report:

the optimal quotient X/E★,

the objective value J(E★),

the distortion Dπ(E★),

the complexity K(E★),

the constraints satisfied,

the active constraints determining the optimum,

alternative minimizers if they exist,

and nearby refinements or coarsenings with similar cost.

This information reveals whether the selected ontology is robust.

A large objective gap between E★ and all alternatives indicates a strongly selected object system.

A small gap indicates ontological ambiguity.

Several equal minimizers indicate that the task and constraints do not determine one unique identity policy.

This is not necessarily a failure.

It may reveal a genuine symmetry or underdetermination in the problem.

The search space of identity policies therefore provides more than a location for minimization.

It reveals the architecture of possible objecthood.

Its lattice order distinguishes refinement from coarsening.

Its quotient maps describe abstraction.

Its feasible subsets encode admissibility.

Its objective landscape ranks effective ontologies.

Its minima identify candidate object systems.

The next section defines the distortion term in detail. It asks how the damage caused by quotienting can be measured relative to prediction, dynamics, causality, intervention, and explanation.

3. Task-Relevant Distortion

Every quotient forgets differences.

If x and y belong to one equivalence class under E, the quotient represents them by the same object [x]E. Any distinction between x and y that cannot be recovered from the quotient has been erased.

Some erased differences are irrelevant.

Others are essential.

A change in illumination may be irrelevant to object identity in an image-recognition task. A small molecular difference may be decisive for drug response. Two microscopic configurations may be thermodynamically interchangeable but dynamically distinct over short time scales. Two records may refer to the same person while containing different transaction histories. Two scientific models may make identical predictions in one regime and incompatible predictions in another.

The quality of an identity policy therefore cannot be determined solely by how much it compresses the state space. One must measure what the compression destroys.

This loss is called task-relevant distortion.

Distortion is not an intrinsic property of an equivalence relation alone. It depends on what the quotient is intended to preserve.

The same policy E may have zero distortion for one task and severe distortion for another.

A quotient that preserves temperature and pressure may be excellent for equilibrium thermodynamics but inadequate for molecular reaction dynamics.

A classification that preserves diagnostic outcome may be useful for screening but unsuitable for treatment selection.

An image representation invariant to color may be beneficial for shape recognition but destructive for identifying traffic signals.

Thus distortion is indexed by a task, policy, observable family, transformation regime, or explanatory purpose π.

Write

$D_\pi(E)$

for the distortion of identity policy E relative to π.

The variational objecthood functional then compares the compression gained through E with the distortion introduced by E.

The central question is:

How much task-relevant variation occurs inside the equivalence classes of E?

If every state in one class has the same task-relevant behavior, the quotient causes no distortion for that task.

If equivalent states differ substantially in relevant behavior, the quotient is poor.

This section develops several forms of distortion. It begins with observable preservation, then extends the concept to prediction, dynamics, intervention, causality, and explanation. It concludes by distinguishing worst-case from expected distortion and by characterizing exact zero-distortion quotients.

3.1 Why Erased Differences Must Be Measured

Let E be an equivalence relation on X.

The quotient map

$q_E: X \to X/E$

replaces every state x with its class [x]E.

Once this replacement is made, the quotient cannot distinguish states belonging to the same class.

Any function defined entirely on X/E must assign one value to the whole class.

Suppose a task requires a quantity h(x).

If h varies inside an E-class, no quotient-level function can reproduce h exactly.

The loss is not a defect of the implementation.

It is a mathematical consequence of the chosen identity policy.

The quotient has declared states with different h-values to be one object.

This gives a fundamental principle:

Every identity declaration is also an information-deletion declaration.

To assert

$x \, E \, y$

is to assert that the distinction between x and y may be ignored.

The assertion is valid only relative to a purpose.

Without a task, observable, or transformation regime, there is no general numerical measure of whether the distinction matters.

This is why complexity alone cannot determine objecthood.

The universal relation achieves maximal compression:

all states become one object.

It also deletes every difference in X.

Whether that deletion is acceptable depends entirely on the intended use.

The equality relation deletes nothing.

Its distortion is zero for every task.

It also produces no abstraction.

The variational problem exists because useful object systems normally require selective forgetting.

A distortion functional should satisfy several intuitive properties.

It should be nonnegative:

$D\pi(E) \geq 0$.

Zero should indicate exact preservation of the relevant structure:

$D\pi(E) = 0$

when the quotient loses nothing required by π.

Coarsening should ordinarily not reduce distortion when the task-relevant structure is fixed.

If $E \preceq F$, then every F-class is a union of E-classes. More variation can occur inside the larger F-classes. For many natural distortion measures,

$D\pi(E) \leq D\pi(F)$.

This monotonicity is desirable but not universal. Some task objectives may improve under coarsening because noise or nuisance variation is removed. In such cases the distortion functional must distinguish harmful variation from nuisance variation rather than treating all variation as relevant.

The functional should also be insensitive to relabeling of quotient classes.

Two partitions differing only in class names represent the same identity policy and should have the same distortion.

When possible, distortion should decompose over classes or states so that it can be estimated computationally.

A generic classwise form is

$D_\pi(E) = \text{Aggregate}\{\delta_\pi(C) : C \in X/E\}$,

where $\delta_\pi(C)$ measures the relevant variation inside class C.

The aggregate may be a maximum, sum, expectation, weighted average, or another risk measure.

This form makes the central object of analysis explicit:

the internal heterogeneity of quotient classes.

A class with no relevant internal variation behaves as one object.

A class with large internal variation is a destructive fusion.

Distortion therefore supplies a quantitative refinement of stable quotienting.

Transformation invariance asks whether equivalent states remain equivalent after transformation.

Distortion asks whether treating them as equivalent damages the task even when formal invariance holds.

A relation may be perfectly invariant and still be uselessly coarse.

The universal relation is the simplest example.

Thus structural admissibility and task fidelity are distinct.

The feasible region enforces structural laws.

The distortion functional evaluates practical or scientific adequacy.

3.2 Observables and Target Quantities

The simplest task-relevant structure is an observable

$h: X \to Y$.

The codomain Y may be a set, metric space, probability space, vector space, category, or another output structure.

The observable assigns to every state the quantity that must be predicted, explained, preserved, or communicated.

Examples include:

a class label,

a future outcome,

a physical measurement,

a control action,

a causal response,

a numerical parameter,

or a scientific explanation type.

The identity policy E preserves h exactly when h is constant on every E-class.

Equivalently,

x E y implies h(x) = h(y).

This means that E is contained in the kernel relation of h:

$E \subseteq \mathrm{Ker}(h)$,

where

$\mathrm{Ker}(h) = \{(x, y) \in X \times X : h(x) = h(y)\}$.

When this condition holds, h factors uniquely through the quotient:

$h = \bar{h} \circ q_E$.

The induced observable

$\bar{h}: X/E \to Y$

assigns one value to each quotient object.

The quotient retains exactly enough information to determine h.

If E is not contained in Ker(h), some equivalence class contains states with different h-values.

Exact factorization is impossible.

Distortion measures the severity of this failure.

Suppose Y carries a distance or loss function

$d_Y: Y \times Y \to [0, \infty]$.

The within-class observable variation may be measured by the diameter

$\mathrm{diam}_h(C) = \sup \{d_Y(h(x), h(y)) : x, y \in C\}$.

A worst-class distortion is then

$D_{h,\max}(E) = \sup \{\operatorname{diam}_h(C) : C \in X/E\}$.

Equivalently,

$D_{h,\max}(E) = \sup \{d_Y(h(x), h(y)) : x \, E \, y\}$.

This quantity is zero exactly when h is constant on every E-class.

One may also choose a representative prediction for each class.

Let

$a: X/E \to Y$

assign a quotient-level approximation a(C) to class C.

The statewise loss is

$\ell_{h,a}(x) = d_Y(h(x), a([x]E))$.

The best achievable distortion for policy E is obtained by optimizing over quotient-level assignments:

$D_h(E) = \inf$ over $a: X/E \to Y$ of Aggregate $\{d_Y(h(x), a([x]E))\}$.

This formulation is important.

The quotient policy determines which states must share one prediction.

The decoder a chooses the best common prediction for each class.

For squared Euclidean loss, the optimal class prediction is often the class mean.

For absolute loss, it is often a median.

For classification loss, it may be the most probable label.

Thus distortion depends both on the partition and on the best quotient-level summary assigned to each class.

When several observables must be preserved, write

$H = \{h_\alpha : X \to Y_\alpha\}$.

One may define a combined distortion

$D_H(E) = \Sigma_\alpha \, w_\alpha \, D_{h_\alpha}(E)$,

where $w_\alpha \geq 0$ expresses the importance of each observable.

Alternatively, one may require exact preservation of some observables and penalize approximation of others.

The family H defines the effective purpose of the quotient.

Two states may be merged only when their differences are negligible across the required targets.

If all observables must be preserved exactly, the coarsest exact identity policy is

$E_H = \bigcap_\alpha \mathrm{Ker}(h_\alpha)$.

Two states are equivalent under E_H exactly when every required observable assigns them the same value.

This relation is the maximal compression available without observable distortion.

Additional transformation or descent constraints may require further refinement.

Observables therefore provide the simplest bridge between abstract equivalence and operational objecthood.

They say what the objects must allow us to know.

3.3 Predictive Distortion

Prediction is one of the most important reasons to preserve distinctions.

Let the present state be $x \in X$.

Let Y represent a future outcome.

Suppose each state determines either a future value

$p(x) \in Y$

or a predictive distribution

$P(\cdot \mid x)$.

An identity policy E is predictively exact when equivalent states have identical predictive consequences.

For deterministic prediction,

$x \, E \, y$ implies $p(x) = p(y)$.

For probabilistic prediction,

x E y implies P(· | x) = P(· | y).

Under these conditions, the quotient object [x]E is a sufficient predictor of the future.

The detailed representative is unnecessary.

Predictive distortion measures the disagreement among future predictions inside each equivalence class.

For deterministic outcomes with metric dY, define

Dpred,max(E) =
sup {dY(p(x), p(y)) : x E y}.

For predictive distributions, let δ be a divergence or probability metric. Examples include total variation distance, Wasserstein distance, Hellinger distance, Jensen–Shannon divergence, or task-specific decision loss.

Then define

Dpred,max(E) =
sup {δ(P(· | x), P(· | y)) : x E y}.

A quotient has zero predictive distortion when every state in one object class induces the same future distribution.

This is closely related to predictive state compression and statistical sufficiency.

However, the present interpretation is ontological.

The quotient classes are not merely efficient encodings. They are candidate effective objects because their internal differences have no predictive consequence under the selected horizon and variables.

Predictive distortion depends on time horizon.

Two states may have nearly identical short-term behavior but diverge later.

Let Pt(· | x) be the predictive distribution at time t.

A horizon-weighted distortion may be defined by

Dpred(E) =
Σt wt sup {δ(Pt(· | x), Pt(· | y)) : x E y},

where wt ≥ 0 assigns importance to different future times.

A finite-horizon policy may be coarser than an infinite-horizon policy.

States interchangeable for the next second may not be interchangeable for the next year.

Thus objecthood can depend on temporal scale.

Predictive distortion may also be action-dependent.

If an agent can choose actions $a \in A$, future outcomes depend on both x and a.

Write

$P(\cdot \mid x, a)$.

An identity policy suitable for decision-making should preserve predictions under every relevant action:

$x \mathrel{E} y$ implies $P(\cdot \mid x, a) = P(\cdot \mid y, a)$

for every $a \in A$.

Approximate predictive distortion becomes

$$D_{pred}(E) = \sup \{\delta(P(\cdot \mid x, a), P(\cdot \mid y, a)) : x \mathrel{E} y \text{ and } a \in A\}.$$

This condition is stronger than passive prediction.

Two states may look identical under observed behavior but respond differently when acted upon.

A policy that merges them may predict passive data accurately while producing incorrect decisions.

The action-conditioned formulation anticipates causal and intervention distortion.

Predictive distortion also distinguishes useful compression from arbitrary clustering.

A cluster may be geometrically compact but predictively heterogeneous.

Another may be geometrically broad but predictively uniform.

For object formation, the second may be superior.

The relevant geometry is induced by future consequences, not necessarily by raw state similarity.

One may define a predictive pseudometric on X:

$d_{pred}(x, y) = \delta(P(\cdot \mid x), P(\cdot \mid y))$.

Candidate object classes should have small diameter under this pseudometric.

Exact predictive equivalence is the zero-distance relation:

x Epred y exactly when dpred(x, y) = 0.

Approximate objecthood permits classes of bounded predictive diameter.

The parameter controlling acceptable distortion determines the resolution of the predictive ontology.

3.4 Dynamical Distortion

Predictive distortion concerns outcomes. Dynamical distortion concerns the evolution of the state or quotient itself.

Let a transformation monoid G act on X.

For g ∈ G, the state x evolves or transforms to g(x).

An equivalence relation E is exactly invariant when

x E y implies g(x) E g(y)

for every g ∈ G.

This condition guarantees that every transformation induces a well-defined action on X/E.

Exact invariance is a hard structural condition.

Approximate systems may require a graded measure of failure.

Suppose X or an observable representation of X carries a distance d.

A one-step dynamical distortion may be defined by measuring how differently equivalent states evolve:

Ddyn,1(E) =
sup {dE(g(x), g(y)) : x E y and g ∈ G0},

where G0 is a selected generating family and dE measures separation relevant to the quotient.

If dE is zero exactly when two states remain E-equivalent, then zero distortion recovers exact invariance.

For a finite quotient, one may use a binary defect:

dE(u, v) =
0 if u E v,

1 otherwise.

Then

$D_{dyn,1}(E) = 0$

exactly when E is invariant under every tested transformation.

A more informative measure uses task-relevant observable distance after transformation.

Let $h: X \to Y$. Define

$$D_{dyn,h}(E) = \sup \{d_Y(h(g(x)), h(g(y))) : x \, E \, y \text{ and } g \in G\}.$$

This asks whether equivalent states remain behaviorally similar after every admissible transformation.

For time-indexed dynamics g_t, define a trajectory distortion

$$D_{traj}(E) = \sup \{\Sigma_t \, w_t \, d_Y(h(g_t(x)), h(g_t(y))) : x \, E \, y\}.$$

Continuous-time systems may replace the sum with an integral over time.

This formulation captures autonomous macro-dynamics.

A quotient object is dynamically effective when states within one class generate approximately the same quotient-level trajectory.

If internal differences grow and produce divergent observable behavior, the class is unstable.

Dynamical distortion can be directional.

Two states may remain close under forward evolution but differ under reversal.

Dissipative systems often erase microscopic distinctions over time. A coarse quotient may become more accurate after a transient period.

One may therefore define delayed distortion:

$$D_{dyn,T}(E) = \sup \{d_Y(h(g_t(x)), h(g_t(y))) : x \, E \, y \text{ and } t \geq T\}.$$

This measures whether states become effectively interchangeable after relaxation time T.

Such a quotient may describe asymptotic objects even when it is inaccurate initially.

Conversely, two states may agree for long periods and diverge only after instability.

A finite-horizon quotient may be useful although it is not globally invariant.

The transformation family G must therefore be selected explicitly.

Object stability is always relative to the transformations and time scales that matter.

Dynamical distortion also applies to stochastic systems.

Let K(x, ·) be a transition kernel.

Equivalent states should induce the same distribution over quotient objects.

Let qE push the transition distribution forward to X/E. Denote the quotient transition law by

$\bar{K}E(x, \cdot) = qE * K(x, \cdot)$.

Exact stochastic stability requires

x E y implies $\bar{K}E(x, \cdot) = \bar{K}E(y, \cdot)$.

Approximate distortion may be measured by

$$D_{stoch}(E) = \sup \{\delta(\bar{K}E(x, \cdot), \bar{K}E(y, \cdot)) : x \, E \, y\}.$$

This is closely related to lumpability and probabilistic bisimulation.

The quotient behaves as a Markovian object system only when internal representatives induce the same quotient-level transition probabilities.

Thus dynamical distortion measures whether the proposed objects possess autonomous evolution.

A low-distortion quotient supports a closed effective theory.

A high-distortion quotient requires hidden state information to predict its own dynamics.

Such a quotient is not truly autonomous, even if its classes appear visually or statistically coherent.

3.5 Causal and Intervention Distortion

Prediction alone does not guarantee causal adequacy.

Two states may produce the same observed outcome under the current environment while responding differently to intervention.

If they are merged, the quotient may predict passive observations correctly but fail to support control, treatment, or causal explanation.

Let A be a set of interventions.

For each intervention $a \in A$, let

$r(a, x)$

denote the response of state x.

The response may be deterministic or probabilistic.

An identity policy E has zero intervention distortion when

$x \, E \, y$ implies $r(a, x) = r(a, y)$

for every $a \in A$.

Equivalent states then have identical intervention profiles.

The quotient class contains all information required to select or evaluate interventions.

For deterministic responses with metric d_R, define

$$D_{int}(E) = \sup \{d_R(r(a, x), r(a, y)) : x \, E \, y \text{ and } a \in A\}.$$

For probabilistic responses, replace d_R with a divergence between response distributions.

The intervention profile of a state can be represented as the function

$R_x: A \to R$

given by

$R_x(a) = r(a, x)$.

Two states are causally equivalent relative to A when their intervention profiles are identical.

The exact causal identity relation is therefore

$E_{causal} = \mathrm{Ker}(x \mapsto R_x)$.

A quotient coarser than E_{causal} merges states with different intervention behavior.

Such a quotient may be predictively adequate under passive data yet causally distorted.

This distinction is central in medicine.

Two patients may have similar observed symptoms and short-term prognosis but respond differently to treatment.

A predictive classification may merge them.

A treatment-selection ontology must separate them.

The same issue appears in engineering.

Two system states may have identical outputs under current inputs but respond differently to control actions.

A quotient used for monitoring may merge them.

A quotient used for control must preserve the difference.

Causal distortion also depends on the intervention set.

If only a restricted family of actions is available, states need be equivalent only under those actions.

Adding new interventions may refine the optimal ontology.

Scientific experimentation therefore changes the space of objecthood.

A new intervention can reveal that a previously unified object class contains causally distinct states.

Object splitting is then not arbitrary. It is forced by newly accessible counterfactual structure.

Causal equivalence may also require preservation of relations among variables rather than only responses.

Suppose a causal model assigns to each state a structural mechanism Mx.

Two states may be considered equivalent when their relevant mechanisms are isomorphic or produce the same interventional distributions.

A causal distortion may compare mechanisms directly or compare their consequences under a selected intervention family.

This provides a route from quotient objecthood to mechanistic explanation.

The intervention formulation also clarifies admissibility.

To treat x and y as one object is to claim that replacing one by the other does not alter any licensed intervention outcome.

Identity is therefore a substitution rule under action, not merely a descriptive resemblance.

3.6 Explanatory Distortion

Explanation is more difficult to quantify than prediction or intervention because explanatory adequacy may involve structure, intelligibility, mechanism, and compression simultaneously.

Nevertheless, the same quotient principle applies.

An identity policy has low explanatory distortion when one explanation can apply coherently to every state in an equivalence class.

Let

$e: X \to Z$

assign to each state an explanatory description, mechanism type, proof pattern, causal model, or theoretical role.

If

$x \, E \, y$ implies $e(x) = e(y)$,

then the explanation factors through the quotient.

Each object class has one explanation.

If e varies inside a class, the quotient hides explanatory differences.

A basic explanatory distortion is therefore

$$D_{exp}(E) = \sup \{d_Z(e(x), e(y)) : x \, E \, y\},$$

where d_Z measures difference among explanations.

The challenge is defining d_Z.

For symbolic explanations, one may use structural edit distance, logical inconsistency, proof difference, or model disagreement.

For mechanistic explanations, one may compare causal graphs, parameterized mechanisms, or intervention profiles.

For scientific theories, one may compare residual error after one common model is fitted to a class.

A model-based formulation is often more useful.

Let M be a family of explanatory models.

For each quotient class C, choose one model mC ∈ M.

Let

Lexp(x, mC)

measure how poorly mC explains state x.

Then define

Dexp(E) =
inf over class models {mC}
of Aggregate {Lexp(x, m[x]E) : x ∈ X}.

This asks how well one common explanatory model can account for all states in each class.

A class is explanatory coherent when one compact model explains its members without large residuals.

A destructive fusion forces unlike mechanisms into one object and raises explanatory loss.

The model-based formulation creates a close relationship between object formation and theory formation.

Selecting an equivalence relation determines which cases should share one explanation.

Selecting class models determines what those explanations are.

The joint optimization becomes

minimize over E and {mC}

Aggregate explanatory loss
+
complexity of the quotient
+
complexity of the class models.

This resembles minimum description length, but the present emphasis is on identity.

The system is not only choosing a concise theory. It is choosing which cases count as instances of the same explanatory object.

Explanatory distortion may differ from predictive distortion.

Two models can make similar predictions while invoking different mechanisms.

A quotient optimized only for prediction may merge mechanistically distinct states.

This may be acceptable for forecasting but unacceptable for scientific understanding.

Conversely, two states may share one mechanism while differing in noisy outcomes. A mechanism-based quotient may merge them even when their individual predictions differ slightly.

Thus predictive and explanatory objecthood can generate different partitions.

The policy π must state which kind of preservation matters.

Explanatory distortion can also be local-to-global.

Different observers may possess local explanations for traces of one global object.

A global class has low explanatory distortion when those local explanations can be integrated into one coherent account or related through lawful translations.

If the explanations contradict one another or require incompatible identities, the global fusion is explanatorily defective.

This suggests a future descent theory for explanations themselves.

Local object classes would carry explanatory models.

Overlap maps would compare them.

A global explanation would be a descent object satisfying coherence conditions.

The current scalar distortion framework provides only the first step.

3.7 Worst-Case and Expected Distortion

A distortion functional must specify how local losses are aggregated.

Two major approaches are worst-case distortion and expected distortion.

Worst-case distortion asks for the largest error produced by the quotient.

For an observable h,

$D_{h,\max}(E) = \sup \{d_Y(h(x), h(y)) : x \mathrel{E} y\}.$

This measure is conservative.

One severe internal disagreement makes the entire policy costly.

Worst-case distortion is appropriate when rare failures are unacceptable.

Safety-critical control, formal verification, legal identity, and exact mathematical abstraction often require this standard.

Its advantages are clarity and robustness.

Its disadvantage is sensitivity to outliers and noise.

One anomalous state can prevent an otherwise useful merger.

Expected distortion weights losses by a probability distribution μ on X or on state pairs.

A representative-based form is

$$D_{h,\mathrm{avg}}(E) = \inf_{a: X/E \to Y} E_\mu[d_Y(h(x), a([x]_E))].$$

A pairwise form is

$$D_{h,\mathrm{pair}}(E) = E[d_Y(h(x), h(y)) \mid x \mathrel{E} y],$$

with an appropriate distribution over equivalent pairs.

Expected distortion is suitable when average performance matters and rare deviations can be tolerated.

It connects naturally to rate-distortion theory and statistical learning.

Its weakness is that severe errors can be hidden by low probability.

A hybrid approach uses risk-sensitive aggregation.

One may combine mean and maximum:

$$D_{\mathrm{hybrid}}(E) = \alpha D_{\mathrm{avg}}(E) + (1 - \alpha) D_{\max}(E).$$

One may use a high quantile of the loss distribution.

One may use conditional value at risk to emphasize the worst tail.

One may impose a hard maximum bound while minimizing expected loss.

The choice reflects the application's tolerance for failure.

Distortion may also be weighted by state importance.

Let $w(x) \geq 0$.

States with high scientific, moral, safety, or operational importance contribute more heavily.

This prevents an optimization dominated by frequent but low-consequence cases from sacrificing rare critical distinctions.

Similarly, interventions, time horizons, and observables may carry weights.

A general expected distortion can be written schematically as

$D\pi(E) =$
$E\pi[L\pi(x, [x]E)],$

where the distribution and loss are determined by policy π.

The policy π therefore includes more than the prediction target.

It specifies whose errors matter, which conditions are likely, which interventions are relevant, and how severe different failures are.

This makes objecthood explicitly normative or task-indexed where appropriate.

There is no contradiction in the same states belonging to different optimal object systems under different π.

A safety policy may preserve distinctions that a compression policy ignores.

A causal policy may refine a predictive policy.

A short-horizon policy may coarsen a long-horizon policy.

A worst-case policy may be finer than an average-case policy.

These are not competing claims about one context-free ontology. They are different admissibility regimes.

However, some policies may dominate others across a broad family of tasks.

An identity relation that achieves strong compression while preserving many independent observables, interventions, and transformations may represent a robust object structure.

Such robustness can be measured by evaluating distortion across a family Π of policies:

$$D_{robust}(E) = \sup \{D_\pi(E) : \pi \in \Pi\}$$

or

$$D_{robust}(E) = \sum_{\pi \in \Pi} w_\pi D_\pi(E).$$

A robust object survives changes in task.

This may help distinguish merely convenient classifications from deeply stable objects.

3.8 Zero-Distortion Quotients

Zero distortion is the exact limit of the theory.

An equivalence relation E has zero distortion relative to observable h when

$x \, E \, y$ implies $h(x) = h(y)$.

Equivalently,

$E \subseteq \mathrm{Ker}(h)$.

The observable factors uniquely through X/E.

For a family of observables H,

$D_H(E) = 0$

exactly when

$E \subseteq \bigcap_{h \in H} \mathrm{Ker}(h)$.

The relation

$E_H = \bigcap_{h \in H} \mathrm{Ker}(h)$

is the coarsest identity policy preserving every observable in H exactly.

Any finer policy also preserves them but achieves less compression.

Thus E_H is the exact least-complexity solution when quotient complexity is monotone under refinement and no additional constraints intervene.

The result is simple but fundamental.

Exact object discovery reduces to finding the common kernel of everything that must be preserved.

The same pattern applies to prediction.

Let $P(\cdot \mid x)$ be the predictive distribution associated with x.

Define

x Epred y

exactly when

$P(\cdot \mid x) = P(\cdot \mid y)$.

Then Epred is the coarsest zero-predictive-distortion relation.

For action-conditioned prediction,

x Eact y

exactly when

$P(\cdot \mid x, a) = P(\cdot \mid y, a)$

for every a.

For intervention response,

x Eint y

exactly when

$r(a, x) = r(a, y)$

for every intervention a.

For explanation maps,

x Eexp y

exactly when

$e(x) = e(y)$.

For several requirements, take the intersection:

Ezero =
Epred ∩ Eint ∩ Eexp ∩ ⋯.

This relation preserves all designated structures exactly.

Transformation invariance must still be checked.

If Ezero is not invariant under G, further refinement may be required.

Similarly, distributed closure coherence may require restricting the feasible family.

The exact variational problem is therefore:

Find the coarsest E such that

$D\pi(E) = 0$,

$E \in Inv(G)$,

$C\mathcal{U}(E) = E$,

and all protected distinctions are preserved.

If such a greatest feasible relation exists, it defines the maximally compressed zero-distortion object system.

It may fail to exist uniquely.

Two incomparable zero-distortion policies may satisfy different structural constraints while having equal complexity.

The feasible zero-distortion family may have several maximal elements.

Additional tie-breaking principles are then required.

Zero distortion should not always be demanded.

Noise, finite data, model uncertainty, and continuous variation often make exact preservation impossible or undesirable.

A zero-distortion quotient may remain so fine that it provides little abstraction.

Approximate objecthood permits controlled loss.

The central variational problem then becomes

$E\bigstar \in \arg\min \{D\pi(E) + \lambda K(E) : E \text{ is feasible}\}$.

The zero-distortion relation provides one endpoint of the compression-distortion frontier.

As the tolerated distortion increases, coarser policies become available.

The quotient ontology changes accordingly.

The central conclusions of this section are therefore:

Distortion measures task-relevant variation inside proposed object classes.

It is indexed by the observable, prediction, transformation, intervention, explanation, or policy that must be preserved.

Exact preservation is equivalent to factorization through the quotient.

Predictive distortion measures disagreement in future consequences.

Dynamical distortion measures failure of autonomous quotient evolution.

Intervention distortion measures disagreement under action.

Explanatory distortion measures failure of one coherent model to apply across a class.

Worst-case and expected distortion represent different tolerances for error.

Zero-distortion policies are contained in the kernels of all required structures.

The next section develops the second half of the variational objective. It asks how the complexity and compression of quotient object systems should be measured and why class count alone is not always sufficient.

4. Complexity and Compression of Object Systems

Distortion measures what a quotient destroys.

Complexity measures what the quotient retains and how difficult the resulting object system is to represent, describe, compute, or use.

The two quantities play opposite roles in the variational principle.

A finer identity policy usually preserves more information but produces a more complex quotient.

A coarser policy usually produces a simpler quotient but permits more variation inside each object class.

The objecthood functional

$J_{\pi,\lambda}(E) = D_{\pi}(E) + \lambda K(E)$

therefore requires a meaningful definition of $K(E)$.

The simplest possibility is to count quotient objects.

If X is finite and E has k equivalence classes, then

$K(E) = k$

or

$K(E) = \log k$.

This captures one important form of compression. A quotient with ten classes is simpler in cardinality than a quotient with one thousand classes.

Class count alone is not always sufficient.

Two quotients may have the same number of objects while differing greatly in the complexity of their classes, the rules needed to recognize membership, the structure of their transformations, or the amount of information required to describe the equivalence relation.

A partition into intervals may be simple.

An arbitrary partition with the same number of blocks may require a long lookup table.

A quotient with a small number of classes may possess complicated transition dynamics.

A quotient with many classes may follow a simple geometric or algebraic rule.

Complexity must therefore be indexed by what the object system is expected to support.

Several kinds of complexity can be distinguished.

Cardinal complexity counts quotient objects.

Informational complexity measures the information required to specify one quotient object.

Policy complexity measures the cost of describing or deciding the equivalence relation.

Structural complexity measures relations, transformations, and organization among quotient objects.

Computational complexity measures the resources needed to construct or use the quotient.

Explanatory complexity measures the length or intricacy of the theory required to describe the objects and their behavior.

These quantities need not agree.

A mature variational theory should not assume that one scalar measure is universally correct. It should identify the complexity notion appropriate to the task policy π.

Nevertheless, all useful complexity measures should express a common principle:

An effective object system is valuable partly because it replaces a detailed state description with a smaller, simpler, or more reusable description.

This section develops cardinal, logarithmic, descriptive, structural, and representation-independent measures of quotient complexity. It also examines compression ratios, the degeneracy of maximal compression, and the need to balance simplicity against distortion.

4.1 Quotient Cardinality as Elementary Complexity

Let X be a finite state space with

$|X| = n$.

Let $E \in Eq(X)$ have quotient classes

$C1, C2, \ldots, Ck$.

Then

$|X/E| = k$.

The most elementary complexity measure is

$Kcard(E) = |X/E|$.

This measure counts the number of effective objects produced by the identity policy.

The equality relation has

$Kcard(\Delta X) = |X| = n$.

Every state remains distinct.

The universal relation has

$Kcard(\nabla X) = 1$.

All states become one object.

If $E \preceq F$, then F is coarser than E and

$|X/F| \leq |X/E|$.

Therefore Kcard is monotone with respect to coarsening:

$E \preceq F$ implies $Kcard(F) \leq Kcard(E)$.

This monotonicity makes quotient cardinality a natural measure for variational objecthood.

Every class merger weakly decreases complexity.

Every class split weakly increases complexity.

The measure is also representation-independent in a basic sense. Relabeling states or classes does not change the number of quotient objects.

Kcard is especially useful when every quotient object carries approximately equal cost.

Suppose a classifier must store one prototype per object.

Suppose a symbolic reasoner must maintain one rule state per concept.

Suppose a finite-state machine uses one internal state per equivalence class.

Suppose a database stores one canonical record per entity.

In such cases, the number of quotient objects directly approximates storage, model size, or operational burden.

Cardinality also gives an elementary measure of object formation.

If n detailed states become k quotient objects, the policy has reduced the ontology by

$n - k$

objects.

A normalized reduction is

$1 - k/n$.

This quantity equals zero for the equality policy and approaches one as the quotient becomes small relative to X.

However, class count ignores class size distribution.

Consider two policies on a state space of one hundred states.

One policy has ten classes of ten states each.

Another has one class of ninety-one states and nine singleton classes.

Both have Kcard = 10.

They produce very different forms of compression.

The first distributes compression evenly.

The second collapses nearly the entire state space into one object while preserving nine exceptional states.

Depending on the task, those structures may have different informational or scientific meaning.

Class count also ignores whether the partition has a concise rule.

A policy separating states by one binary property may be simpler to describe than a random partition with the same number of classes.

Thus quotient cardinality is a useful first complexity measure, but not a complete one.

It measures the size of the effective ontology.

It does not measure the complexity of constructing, recognizing, or explaining that ontology.

For infinite state spaces, cardinality becomes less informative.

Two quotients may both have infinitely many classes while differing substantially in dimension, topology, measure, or descriptive structure.

Additional complexity measures are then required.

Even in finite systems, cardinality should usually be understood as one term in a broader complexity functional.

4.2 Logarithmic Object Complexity

When quotient objects function as alternatives that must be identified, logarithmic complexity is often more natural than direct class count.

Define

$Klog(E) = \log |X/E|$.

The logarithm base determines the unit.

Base 2 measures bits.

The natural logarithm measures nats.

Any fixed base produces the same ordering of finite quotients.

The logarithmic form represents the idealized information needed to specify one object when all quotient classes are treated as equally likely.

If the quotient has k objects, an optimal fixed-length binary code requires approximately

$\log_2 k$

bits per object label.

The equality policy therefore has complexity

$K\log(\Delta X) = \log |X|$.

The universal relation has

$K\log(\nabla X) = \log 1 = 0$.

The logarithmic measure remains monotone under coarsening:

$E \preceq F$ implies $K\log(F) \leq K\log(E)$.

Its advantage is diminishing sensitivity to class count.

Reducing a quotient from one thousand classes to one hundred is treated as comparable in logarithmic scale to reducing one hundred classes to ten.

Both changes reduce the number of possible object labels by a factor of ten.

This often matches information-theoretic intuition better than an absolute class reduction.

If quotient classes are not equally probable, entropy provides a more refined measure.

Let μ be a probability distribution on X.

The probability of a quotient class C is

$\mu(C) = \Sigma x \in C\ \mu(x)$

in the finite case.

The quotient entropy is

$H\mu(E) = -\Sigma C \in X/E\ \mu(C) \log \mu(C)$.

This measures the expected information required to identify the quotient object $qE(x)$.

The equality relation gives the entropy of the original state distribution:

$H\mu(\Delta X) = H(\mu)$.

The universal relation gives

$H\mu(\nabla X) = 0$.

Coarsening cannot increase quotient entropy because merging categories removes distinctions.

Thus, if $E \preceq F$,

$H\mu(F) \leq H\mu(E)$.

Entropy accounts for unequal class frequency.

A quotient with many extremely rare objects may have lower expected coding cost than its class count suggests.

A quotient with a few nearly equally probable objects may carry substantial information.

Entropy therefore measures effective rather than merely nominal object complexity.

One may define the information removed by E as

$I_{removed}(E) = H(\mu) - H_\mu(E)$.

This is the reduction in state-label uncertainty achieved by replacing the detailed state with its quotient class.

Because q_E is a deterministic function of x, the quotient entropy is also the mutual information between the state and its quotient label:

$I(x; q_E(x)) = H_\mu(E)$.

The quotient retains $H_\mu(E)$ units of information about which object is present.

It discards the remaining within-class information.

This produces a clean interpretation.

Complexity is the amount of object-level information retained.

Compression is the amount of state-level information discarded.

The distinction is important.

A low-complexity quotient retains little identifying information.

A high-complexity quotient retains more.

The variational principle must decide how much retained information is justified by reduced distortion.

Entropy-based complexity requires a probability distribution.

This may be empirical, theoretical, stationary, or policy-dependent.

Different distributions can rank the same partitions differently.

A quotient optimized for common states may ignore distinctions among rare states.

A worst-case policy may reject that result.

Thus entropy is appropriate when expected coding or average occurrence matters, but it does not replace cardinal or worst-case measures in every application.

The logarithmic family of measures provides a bridge from partition counting to information theory.

It expresses object complexity in units of distinguishability rather than raw class number.

4.3 Description Length of Equivalence Policies

The quotient size does not determine the cost of describing the identity policy itself.

Let L(E) denote the length of a description of E in a selected representation language.

This may be called policy description length.

A partition can be represented extensionally by listing the class label of every state.

For a finite state space of n states and k classes, such a representation may require approximately

n log k

bits, ignoring coding overhead.

This is a lookup-table description.

It treats the partition as arbitrary.

Many meaningful policies have shorter intensional descriptions.

Examples include:

states are equivalent when they have the same temperature,

images are equivalent under rotation,

graphs are equivalent up to isomorphism,

records are equivalent when specified identifiers agree,

and physical states are equivalent when selected conserved quantities coincide.

These policies can be expressed by rules rather than by enumerating every class member.

A simple rule may generate a large and highly compressed quotient.

Thus one should distinguish:

the complexity of the quotient labels,

and the complexity of the policy defining them.

A general description-length complexity may take the form

$$K_{desc}(E) = L(E) + L(X/E \mid E).$$

The first term describes the identity rule.

The second describes any additional structure required to specify or use the resulting quotient.

In some settings, E determines the quotient completely, so the second term is negligible.

In others, quotient objects carry prototypes, parameters, names, or models that must also be encoded.

A full two-part description may be

$$K_{total}(E) = L(E) + \Sigma_{C \in X/E} L(m_C) + L(\text{residual data} \mid E, \{m_C\}),$$

where m_C is a model or prototype associated with class C.

The residual term may overlap conceptually with distortion.

The exact division between model complexity and distortion depends on the coding framework.

Description length supports an important distinction between simple and arbitrary partitions.

Consider a state space of integers from 1 to one million.

The policy “equivalent when parity agrees” has two classes and a very short rule.

An arbitrary two-class partition has the same quotient cardinality but may require nearly one million binary assignments to describe.

Both have two objects.

Their identity policies differ greatly in structural simplicity.

A scientific object system is often valued because its classes can be recognized through concise laws.

Object complexity should therefore include the rule of identity, not merely the number of objects.

Description length is language-dependent.

A relation that is simple in one representation may be complicated in another.

A coordinate change can transform a simple threshold into a complex boundary.

A mature theory must either select a privileged description language or seek invariance across an admissible family of representations.

This issue cannot be eliminated entirely.

Complexity is always relative to available primitives.

A physical law is concise relative to a mathematical language containing appropriate operations.

A biological category is concise relative to a vocabulary of traits or ancestry.

An AI representation is concise relative to its architecture.

The notation π can therefore include the representation language used to evaluate complexity.

One may write

$K_\pi(E)$

rather than K(E) when this dependence matters.

A related computational measure is the cost of deciding equivalence.

Let Tdec(E) be the resources required to determine whether

x E y.

A policy may have few classes but a difficult membership test.

Graph isomorphism illustrates the general form of the issue: the quotient concept may be concise while equivalence checking is computationally demanding.

Thus practical object complexity may include both description length and decision complexity.

A quotient ontology that cannot be evaluated efficiently may be mathematically elegant but operationally poor.

4.4 Structural Complexity Beyond Class Count

Quotient objects rarely exist as an unstructured list.

They may carry transitions, algebraic operations, causal relations, spatial organization, probabilities, morphisms, or explanatory models.

The complexity of this structure can dominate the complexity of the object count.

Suppose a transformation $g: X \to X$ preserves E and induces

$\bar{g}: X/E \to X/E$.

The quotient dynamics may be represented by a directed graph.

The nodes are quotient objects.

An edge $C \to D$ is present when $\bar{g}(C) = D$ or when a transition from C to D is possible.

Two quotients with the same number of classes can have very different transition complexity.

One may form a simple cycle.

Another may form a dense graph with many transition types.

Define a dynamical complexity measure such as

$K_{dyn}(E)$ = number of quotient transitions,

or a weighted version that includes transition probabilities, labels, or parameters.

For a family G of transformations, one may count the distinct induced maps

$\{\bar{g} : g \in G\}$.

A quotient that reduces both state count and transformation complexity may provide a stronger abstraction than one that reduces state count alone.

Algebraic object systems have analogous structure.

If X carries operations and E is a congruence, then X/E inherits quotient operations.

The complexity of the quotient algebra depends on:

the number of elements or classes,

the number and arity of operations,

the complexity of operation tables,

the number of generators,

the number of defining relations,

or the length of a presentation.

A quotient with many elements may have a short algebraic presentation.

A smaller quotient may require an irregular operation table.

Geometric quotient objects may be evaluated by dimension, curvature description, topological invariants, singularities, or atlas complexity.

Probabilistic quotients may be evaluated by transition entropy, parameter count, or model rank.

Categorical quotients may be evaluated by numbers of objects and morphisms, generators and relations, or complexity of composition.

Thus a general structural complexity may be written

Kstruct(E) =
Kobjects(E)
+
Krelations(E)
+
Koperations(E)
+
Kdynamics(E).

The weights assigned to these terms depend on the intended use.

For prediction, quotient transition complexity may dominate.

For storage, object count may dominate.

For explanation, description length of laws may dominate.

For control, action-response structure may dominate.

A quotient can even increase structural complexity.

Merging states can produce complicated effective dynamics because hidden internal differences must be represented probabilistically or non-Markovianly.

A coarse quotient may have fewer objects but require memory of past states to predict future behavior.

The apparent reduction in state complexity is then offset by increased dynamical complexity.

This is a crucial warning.

Compression of the object set does not guarantee compression of the theory.

Suppose E merges states that have slightly different transition behavior.

The quotient may no longer support a deterministic map.

It may require a stochastic kernel.

If transition probabilities depend on hidden histories, it may require a higher-order model.

The quotient object count decreases, but the effective dynamics become more complex.

A comprehensive variational theory should include this induced complexity.

One possible functional is

$$K_{eff}(E) = \alpha \log |X/E| + \beta L(E) + \gamma L(\text{dynamics on } X/E),$$

where α, β, and γ are nonnegative weights.

The preferred quotient then minimizes the total complexity of both objects and laws.

This formulation aligns with scientific abstraction.

A useful scientific theory does not merely name fewer entities.

It also expresses their behavior through simpler relations.

The deepest compression may occur when many detailed states collapse into a small object space governed by concise dynamics.

4.5 Compression Ratios

A compression ratio compares the complexity of the quotient with the complexity of the original state description.

For finite cardinal complexity, define

$$R_{card}(E) = |X/E| / |X|.$$

The ratio lies between

$$1/|X|$$

and

1.

The equality relation gives

$R_{card}(\Delta X) = 1$.

The universal relation gives

$R_{card}(\nabla X) = 1/|X|$.

A related compression gain is

$G_{card}(E) = 1 - |X/E|/|X|$.

This equals zero under equality and approaches one under strong compression.

Cardinality ratios are intuitive but do not express coding cost linearly.

A logarithmic ratio may be defined as

$R_{log}(E) = \log |X/E| / \log |X|$

when $|X| > 1$.

The equality policy gives $R_{log} = 1$.

The universal policy gives $R_{log} = 0$.

This ratio measures the fraction of idealized object-label information retained.

An entropy-based ratio is

$R_{\mu}(E) = H_{\mu}(E) / H(\mu)$

when $H(\mu) > 0$.

The corresponding information-loss fraction is

$1 - H_{\mu}(E)/H(\mu)$.

These ratios account for nonuniform state probabilities.

Description-length compression may be defined as

$R_{desc}(E) =$
compressed description length under E
divided by
uncompressed description length.

The compressed description may include:

the policy description,

the quotient object label,

class models,

and residual data.

A ratio below one indicates compression.

A ratio above one indicates that the quotient representation costs more than the original description.

This can happen.

A complicated partition may introduce overhead without producing enough regularity to compensate.

Thus not every quotient is an effective compression in descriptive terms.

Compression ratios are useful for comparing systems of different size.

Reducing one million states to one thousand objects and reducing one thousand states to one object both produce a factor of one thousand in cardinality.

An absolute class-count difference would treat them very differently.

Ratios expose relative compression.

However, compression ratio alone still does not evaluate distortion.

A universal quotient maximizes compression regardless of scientific value.

Compression must be placed on a frontier against task loss.

For every E, plot

K(E)

against

Dπ(E).

Policies not dominated by another policy form a compression-distortion frontier.

A policy E is dominated when another policy F satisfies

$K(F) \leq K(E)$

and

$D\pi(F) \leq D\pi(E)$,

with at least one strict inequality.

A dominated policy is never optimal for any positive weighting of the two objectives.

It is both more complex and more distorted than an alternative.

The undominated policies form the meaningful candidate set.

As λ varies, the variational optimum moves along this frontier.

This geometric view will become central in the analysis of phase transitions in conceptual resolution.

Compression ratios also permit comparison with a null or baseline ontology.

The baseline may be equality, an existing human classification, a previously learned policy, or a random partition.

The gain from a new identity policy can then be evaluated relative to what it replaces.

This is often more meaningful than comparison with equality alone.

4.6 Representation-Independent Complexity

An identity policy is a structural object. Its complexity should not change merely because states or classes are renamed.

Let $\sigma: X \to X$ be a bijection.

The transported relation σE is defined by

$\sigma(x)\ \sigma E\ \sigma(y)$

exactly when

$x\ E\ y$.

The policies E and σE have the same partition structure under relabeling.

A representation-independent complexity measure should satisfy

$K(\sigma E) = K(E)$.

Cardinality, class-size multiset, and quotient entropy under correspondingly transported probability measures satisfy this condition.

A raw description length based on state names may not.

Representation invariance can be extended.

Suppose

$f: (X, E) \to (Y, F)$

is an isomorphism of policy-indexed spaces.

Then a structural complexity measure should ideally satisfy

$K_X(E) = K_Y(F)$.

This says that isomorphic object systems have the same complexity.

The requirement is natural but difficult to satisfy for all forms of complexity.

Computational complexity depends on representation.

An equivalence test may be easy in one encoding and difficult in another.

Description length depends on language.

A structure may have a short description in one coordinate system and a long description in another.

Thus complete representation independence may be impossible or undesirable.

The appropriate principle is often invariance under an admissible representation class.

Let R be a family of transformations regarded as mere changes of presentation.

A complexity measure is R-invariant when

$K(rE) = K(E)$

for every $r \in R$.

The family R may include:

state relabelings,

coordinate transformations,

basis changes,

graph isomorphisms,

permutations of features,

or other representation symmetries.

This prevents the variational theory from selecting an ontology merely because of arbitrary notation.

One may also minimize description length over admissible representations:

$K_{inv}(E) = \inf \{L(rE) : r \in R\}$.

This assigns to the policy the shortest description available within the representation family.

The construction is conceptually attractive but may be computationally difficult.

Another approach is to define complexity through invariants.

Examples include:

number of quotient classes,

class-size distribution,

dimension,

rank,

entropy,

numbers of quotient transitions,

numbers of generators,

or topological invariants.

These quantities ignore some representational details by construction.

They also omit potentially important complexity.

No single invariant captures every aspect of an object system.

A layered measure may therefore be preferable.

The first layer is fully structural and representation-independent.

The second layer measures computational or descriptive cost in a chosen implementation.

One might write

$K(E) = K_{struct}(E) + \eta K_{impl}(E)$,

where K_{struct} is invariant under approved isomorphisms and K_{impl} records practical representation cost.

This separation prevents implementation details from being confused with ontology while still acknowledging that practical object systems must be computed.

Representation independence also has philosophical significance.

If an object system changes whenever its states are renamed, it is not detecting structure.

It is detecting notation.

The identity policy should capture relations and invariants that survive admissible changes of presentation.

Variational objecthood therefore requires not only compression, but compression of the right structure.

4.7 Degenerate Maximal Compression

Every unconstrained complexity minimization has a trivial solution.

The universal relation ∇X produces one quotient object and therefore minimizes:

quotient cardinality,

logarithmic class complexity,

quotient entropy,

and many other monotone object-count measures.

If the variational objective were only

minimize $K(E)$,

the answer would always be

$E = \nabla X$.

All states would become one object.

This is degenerate maximal compression.

The degeneracy shows why complexity cannot serve as the sole principle of object formation.

A useful quotient must preserve task-relevant distinctions.

Distortion prevents trivial collapse.

Protected distinctions, transformation stability, descent, and causal constraints provide additional barriers.

The universal relation may still have zero formal invariance defect because every transformed state remains in the sole class.

Thus transformation invariance alone does not exclude it.

Closure coherence may also permit it.

The universal policy is often a fixed point of closure.

The decisive test is whether the one-object quotient preserves what matters.

If the task observable is constant, universal equivalence may in fact be optimal.

This is not a defect.

When no relevant distinction exists, one object is sufficient.

The universal relation is degenerate only when the task requires variation that it erases.

The equality relation represents the opposite degeneracy.

It always achieves zero distortion for any state-dependent task because it merges no distinct states.

If the objective were only

minimize $D_\pi(E)$,

the equality policy would always be optimal.

No abstraction would occur.

The two extremes therefore correspond to one-sided optimization.

Complexity alone collapses everything.

Distortion alone separates everything.

Objecthood appears only through their competition.

There are other possible degeneracies.

A complexity measure based only on class count may favor a small quotient whose class-membership rule is extremely complicated.

A description-length measure may favor a simple rule producing classes with destructive internal variation.

An average distortion measure may sacrifice rare critical states.

A worst-case measure may preserve many classes because of one noisy outlier.

A dynamical measure may select a quotient useless for explanation.

Each scalar objective creates blind spots.

This motivates composite functionals and hard constraints.

A policy should not be considered optimal merely because it has the smallest numerical value under an inadequately specified K or D.

The quantities must correspond to the intended use.

One may also impose nontriviality constraints.

For example,

$1 < |X/E| < |X|$.

Such constraints exclude the two extremes but do not identify a scientifically meaningful interior policy.

They are sometimes useful in exploratory computation, but they are not a substitute for task-relevant distortion.

A stronger solution is to require preservation of selected observables, interventions, or dynamics.

The quotient is then forced to retain every distinction necessary for those structures.

Maximal compression becomes meaningful because it occurs inside a constrained feasible region.

The resulting principle is:

Compress as far as possible, but no farther than lawful substitution permits.

This is the object-forming analogue of preserving sufficient information.

4.8 Why Complexity Must Be Balanced by Distortion

The two fundamental quantities of variational objecthood are now visible.

K(E) measures the burden of the effective ontology.

$D\pi(E)$ measures the damage caused by treating class members as interchangeable.

Neither quantity is sufficient alone.

The combined objective is

$J\pi,\lambda(E) = D\pi(E) + \lambda K(E)$.

The parameter λ expresses the cost of complexity relative to distortion.

When $\lambda = 0$, the objective ignores complexity.

The equality relation or another zero-distortion refinement is favored.

As λ increases, compression becomes more valuable.

Policies with fewer or simpler objects become competitive.

At sufficiently large λ, very coarse policies may dominate despite substantial distortion.

The parameter therefore controls conceptual resolution.

An equivalent constrained formulation is:

minimize $K(E)$

subject to

$D\pi(E) \leq \varepsilon$.

Here ε is the maximum tolerated distortion.

This formulation asks for the simplest quotient meeting a fidelity requirement.

Another formulation is:

minimize $D\pi(E)$

subject to

$K(E) \leq B$.

Here B is a complexity budget.

This asks for the most faithful ontology that can be supported with limited resources.

The penalized and constrained forms are related but need not produce identical solutions in every discrete problem.

The policy lattice is not a smooth convex space.

Several policies may exchange optimality abruptly as λ, ε, or B changes.

These discontinuities are not mathematical nuisances.

They are candidate phase transitions in objecthood.

The balance also determines whether a merger is justified.

Suppose F is obtained from E by merging two quotient classes.

Define

$\Delta D = D\pi(F) - D\pi(E)$,

and

$\Delta K = K(F) - K(E)$.

Normally,

$\Delta D \geq 0$

and

$\Delta K \leq 0$.

The merger improves the objective when

$\Delta D + \lambda \Delta K < 0$.

Because ΔK is negative, this is equivalent to requiring that the complexity reduction outweigh the distortion increase under the selected weighting.

A split is justified by the reverse inequality.

This gives a local decision rule for object fusion and object birth.

The same principle applies globally, although interactions among classes may prevent greedy decisions from reaching the best policy.

The form of K also affects the selected ontology.

Cardinality complexity favors fewer classes.

Entropy complexity favors lower expected object information.

Description length favors regular identity rules.

Dynamical complexity favors quotients with simple induced evolution.

Explanatory complexity favors object systems governed by concise models.

Different choices can produce different optimal partitions.

This is not necessarily a weakness.

It reveals that simplicity has several meanings.

A physically useful object may minimize dynamical complexity.

A communicatively useful concept may minimize coding complexity.

A scientific theory may balance explanatory and predictive complexity.

An AI representation may balance computational, memory, and task complexity.

The policy π should specify which costs count.

A general composite complexity may be written

$$K_\pi(E) = \alpha K_{card}(E) + \beta H_\mu(E) + \gamma L(E) + \delta K_{dyn}(E) + \eta K_{comp}(E),$$

with nonnegative coefficients selected for the application.

The coefficients should not be treated as arbitrary decorations.

They define the economy under which objects are selected.

A change in resource constraints or scientific purpose may alter the optimal ontology.

The balance between distortion and complexity also suggests a criterion for robust objecthood.

A quotient object system is robust when it remains optimal or nearly optimal across a substantial range of λ and across reasonable choices of complexity measure.

A policy selected only for one finely tuned parameter may be fragile.

A policy lying on a broad plateau of low objective value may represent a more stable effective ontology.

This idea will later support the interpretation of conceptual-resolution plateaus as persistent object regimes.

The principal conclusions of this section are:

Quotient cardinality provides the simplest measure of object-system complexity.

Logarithmic class count and quotient entropy measure retained object information.

Policy description length measures the complexity of the identity rule itself.

Structural complexity includes relations, operations, and dynamics among quotient objects.

Compression ratios compare quotient complexity with the original state description.

Meaningful complexity should be invariant under admissible changes of representation.

Complexity minimization alone produces universal-equivalence collapse.

Distortion minimization alone produces equality-level over-refinement.

Effective objecthood arises from balancing the two.

The next section adds distributed coherence to the variational problem. It develops the globalization–restriction operator as a closure operator, characterizes its fixed points, and incorporates descent defects as either hard constraints or penalties in the objecthood functional.

5. Closure Coherence and Descent Constraints

A variational theory of objecthood must evaluate more than the quality of one equivalence relation on one state space. Many object systems are distributed across observers, instruments, databases, models, scales, or local contexts. Each domain forms its own quotient objects. The global ontology must then be assembled from these partial identity systems.

Local optimization alone is not sufficient.

Each observer may discover a policy that is internally simple, predictive, and transformation-stable. The resulting family can still fail globally. Identities accepted in different domains may compose through overlaps and return to one domain as an identification that its local policy rejects.

The obstruction is generated by equivalence closure.

This creates a new variational requirement.

A distributed identity system should not merely minimize local distortion and complexity. It should also remain unchanged when its local identity declarations are globalized and then restricted back to the observers.

This is closure coherence.

Let X be covered by domains

$\mathcal{U}$ = {Ui}i∈I.

Let

$\mathcal{E}$ = (Ei)i∈I

be a family of local identity policies, with

Ei ∈ Eq(Ui).

The family $\mathcal{E}$ determines a generated global equivalence relation

G$\mathcal{U}$($\mathcal{E}$) = EqCl(⋃i Ei).

Restricting this global relation back to each domain gives a new local family

C$\mathcal{U}$($\mathcal{E}$) =
(G$\mathcal{U}$($\mathcal{E}$)|Ui)i∈I.

The operator C$\mathcal{U}$ is the globalization–restriction operator.

It takes locally declared identities, computes every global identity forced by their composition, and returns the consequences visible to each observer.

A distributed policy family is faithfully coherent exactly when

C$\mathcal{U}$($\mathcal{E}$) = $\mathcal{E}$.

Thus coherent distributed object systems are fixed points of a closure operator.

This equation is one of the central structural laws of the theory.

It says that globalization must add no new identity when viewed from any local domain.

The fixed-point condition can be imposed as a hard constraint. It can also be relaxed by assigning a penalty to the difference between $\mathcal{E}$ and C$\mathcal{U}$($\mathcal{E}$). This permits approximate distributed coherence when local classifications are noisy, provisional, learned from finite data, or subject to revision.

The distinction between hard and penalized descent is consequential.

A hard constraint rejects every family that loses even one local distinction after globalization.

A penalty permits some local object fusion when the resulting reduction in complexity or improvement in predictive performance justifies the cost.

The variational theory must therefore decide not only which states become objects, but also which local distinctions have authority when global closure challenges them.

This section defines the globalization–restriction operator, proves its closure properties, characterizes coherent identity systems as fixed points, introduces quantitative descent defects, and develops distributed optimization over observer networks.

5.1 The Globalization–Restriction Operator

Let X be a state space and let

$\mathcal{U}$ = {Ui}i∈I

be a cover of X.

Thus

X = ⋃i Ui.

For every domain Ui, choose an equivalence relation Ei on Ui.

The full local policy family is

$\mathcal{E}$ = (Ei)i∈I.

The natural ambient space of distributed policies is the product lattice

L$\mathcal{U}$ = ∏i∈I Eq(Ui).

An element of L$\mathcal{U}$ assigns one conceptual resolution to each observer.

The order is componentwise refinement.

For two families

$\mathcal{E}$ = (Ei)

and

$\mathcal{F}$ = (Fi),

define

$\mathcal{E} \preceq \mathcal{F}$

when

Ei ⊆ Fi

for every i.

Thus $\mathcal{E}$ is finer than $\mathcal{F}$ in every local domain.

The globalization map is

$G\mathcal{U}$: $L\mathcal{U}$ → Eq(X),

defined by

$G\mathcal{U}(\mathcal{E})$ = EqCl(⋃i Ei).

The union contains every local identity declaration.

Equivalence closure adds every identity forced by reflexivity, symmetry, and transitivity.

The result is the unique smallest global equivalence relation containing every Ei.

The restriction map is

$R\mathcal{U}$: Eq(X) → $L\mathcal{U}$,

defined by

$R\mathcal{U}$(E) = (E|Ui)i∈I.

The composition

$C\mathcal{U}$ = $R\mathcal{U}$ ∘ $G\mathcal{U}$

is the globalization–restriction operator:

$C\mathcal{U}(\mathcal{E})$ =
(G$\mathcal{U}(\mathcal{E})$|Ui)i∈I.

Its ith component is

$C\mathcal{U}$,i($\mathcal{E}$) = G$\mathcal{U}(\mathcal{E})$|Ui.

The operator compares what observer i originally declared with what the complete network forces observer i to accept.

Every original local identity remains present:

Ei ⊆ $C\mathcal{U}$,i($\mathcal{E}$).

This follows because Ei is contained in the generating union and therefore in G$\mathcal{U}(\mathcal{E})$.

The operator can add identities but cannot remove them.

It acts by local coarsening.

If $C\mathcal{U}$,i($\mathcal{E}$) is strictly larger than Ei, then some locally distinct states have become globally equivalent through a path involving other domains.

The operator is therefore not a mere consistency check. It is a completion process.

It computes the least locally visible family containing all consequences of distributed identity composition.

A second composition is also useful.

Given a global relation E ∈ Eq(X), restrict it and globalize again:

G$\mathcal{U}$(R$\mathcal{U}$(E)) = EqCl(⋃i E|Ui).

This relation need not equal E.

A global identity between states never jointly represented in any connected chain of domains may disappear when only local restrictions are retained.

Thus local observation may fail to recover all optional global identities.

However, when E itself is generated by its local restrictions, one has

G$\mathcal{U}$(R$\mathcal{U}$(E)) = E.

The minimal globalizations constructed from local families satisfy this property automatically.

The two compositions therefore express different questions.

C$\mathcal{U}$ = R$\mathcal{U}$G$\mathcal{U}$ asks whether local policies survive globalization.

G$\mathcal{U}$R$\mathcal{U}$ asks whether a global policy can be reconstructed from its local traces.

The present section emphasizes C$\mathcal{U}$ because faithful object formation requires protection of local distinctions.

5.2 Closure Operators on Distributed Policies

The globalization–restriction operator C$\mathcal{U}$ is a closure operator on the product lattice L$\mathcal{U}$.

A closure operator satisfies three properties:

extensivity,

monotonicity,

and idempotence.

Extensivity means

$\mathcal{E} \preceq C\mathcal{U}(\mathcal{E})$.

Every local identity declared by $\mathcal{E}$ remains valid after globalization and restriction.

No accepted equivalence is lost.

Monotonicity means

$\mathcal{E} \preceq \mathcal{F}$ implies $C\mathcal{U}(\mathcal{E}) \preceq C\mathcal{U}(\mathcal{F})$.

If every local policy in $\mathcal{F}$ is at least as coarse as the corresponding policy in $\mathcal{E}$, then the global relation generated by $\mathcal{F}$ contains the global relation generated by $\mathcal{E}$.

Consequently, all restricted global consequences of $\mathcal{E}$ are also consequences of $\mathcal{F}$.

Idempotence means

$C\mathcal{U}(C\mathcal{U}(\mathcal{E})) = C\mathcal{U}(\mathcal{E})$.

Once all globally forced local identities have been added, repeating the procedure adds nothing further.

To see this, let

Eglobal = $G\mathcal{U}(\mathcal{E})$.

The first closure produces the local family

(Eglobal|Ui)i.

Every original Ei is contained in Eglobal|Ui. Therefore globalizing the closed family must contain the original generated relation Eglobal.

Conversely, every relation Eglobal|Ui is contained in Eglobal. Globalizing their union cannot produce a relation larger than Eglobal.

Thus globalizing the closed family reproduces Eglobal exactly.

Restriction then reproduces the same local family.

Idempotence shows that $C\mathcal{U}$ is not an iterative approximation requiring indefinite repetition.

One application computes the complete distributed closure.

This is possible because equivalence closure already includes every finite identity path.

The operator $C\mathcal{U}$ has an important universal property.

For any local family $\mathcal{E}$, $C\mathcal{U}(\mathcal{E})$ is the least closure-coherent family coarser than $\mathcal{E}$.

Suppose $\mathcal{F}$ is a fixed point of $C\mathcal{U}$ and

$\mathcal{E} \preceq \mathcal{F}$.

Then monotonicity gives

$C\mathcal{U}(\mathcal{E}) \preceq C\mathcal{U}(\mathcal{F})$.

Because $\mathcal{F}$ is fixed,

$C\mathcal{U}(\mathcal{F}) = \mathcal{F}$.

Therefore

$C\mathcal{U}(\mathcal{E}) \preceq \mathcal{F}$.

Thus no smaller coherent coarsening of $\mathcal{E}$ exists.

This makes $C\mathcal{U}$ an order-theoretic projection onto the coherent policy families.

It is not a metric projection. It does not minimize a numerical distance unless a compatible distance is chosen.

It is canonical in the refinement order.

It adds only identities forced by the original family and equivalence closure.

The fixed points of any closure operator form a closure system.

Define

Fix($C\mathcal{U}$) =
$\{\mathcal{E} \in L\mathcal{U} : C\mathcal{U}(\mathcal{E}) = \mathcal{E}\}$.

The family Fix($C\mathcal{U}$) is closed under arbitrary meets.

Let $\{\mathcal{E}\alpha\}$ be fixed-point families.

Their componentwise meet is

$\bigwedge\alpha\ \mathcal{E}\alpha =$
$(\bigcap\alpha\ E\alpha,i)i$.

This meet is also closure-coherent.

Thus coherent distributed ontologies possess greatest common refinements.

Their joins inside the fixed-point lattice are obtained by first taking the componentwise join in $L\mathcal{U}$ and then applying $C\mathcal{U}$.

Therefore Fix($C\mathcal{U}$) is itself a complete lattice.

This result is important for variational objecthood.

The feasible space of exact distributed identity systems is not an arbitrary subset. It retains a complete order-theoretic structure.

Optimization can therefore exploit closure, refinement, meet, and coherent join operations.

5.3 Coherent Identity Systems as Fixed Points

A distributed policy family $\mathcal{E}$ is closure-coherent when

$C\mathcal{U}(\mathcal{E}) = \mathcal{E}$.

Componentwise, this means

$G\mathcal{U}(\mathcal{E})|Ui = Ei$

for every i.

This is exactly the faithful descent condition.

The locally declared identity systems are the precise restrictions of their minimal generated global relation.

No observer is forced to merge two of its original local objects.

The fixed-point equation has several equivalent forms.

First, every global equivalence class intersects each Ui in at most one Ei-class.

Second, the canonical map

$\phi i: Ui/Ei \rightarrow X/G\mathcal{U}(\mathcal{E})$

is injective for every i.

Third, no connected component of the local-object correspondence graph contains two distinct local quotient objects from the same domain.

Fourth, every identity path beginning and ending in Ui has endpoints already equivalent under Ei.

Fifth, every locally preserved distinction survives the complete distributed closure.

These formulations describe the same fixed-point property from different viewpoints.

The closure equation is especially suitable for optimization because it can be imposed directly as a constraint:

$\mathcal{E} \in Fix(C\mathcal{U})$.

A variational problem over distributed policies may therefore be written

$\mathcal{E}\bigstar \in \arg\min \{J(\mathcal{E}) : \mathcal{E} \in \mathrm{Fix}(C\mathcal{U})\}$.

Additional transformation, prediction, complexity, or protected-distinction constraints can be added.

The fixed-point condition is stronger than pairwise overlap agreement.

A family may satisfy

$E_i|_{U_i \cap U_j} = E_j|_{U_i \cap U_j}$

for every pair i, j and still fail

$C\mathcal{U}(\mathcal{E}) = \mathcal{E}$.

Pairwise compatibility prevents direct contradiction.

Closure coherence prevents indirect contradiction generated by transitive paths.

This difference is crucial for variational learning.

If local policies are optimized independently and only their overlaps are compared, the combined system may appear consistent while containing a global fusion defect.

The closure operator exposes such defects immediately.

Fixed points also provide a definition of completed distributed concepts.

A provisional local family may contain only direct observations.

Applying $C\mathcal{U}$ adds all identities logically forced by the network.

If the family changes, it was not closed under its own consequences.

If the family remains unchanged, its local ontology is complete relative to the cover and identity declarations.

This gives closure coherence an epistemic interpretation.

A coherent observer network has already incorporated every identity conclusion that follows from its distributed evidence.

It contains no hidden merger waiting to appear upon integration.

5.4 Descent Defects as Constraint Violations

When $\mathcal{E}$ is not a fixed point, the difference

$C\mathcal{U}(\mathcal{E}) - \mathcal{E}$

measures the identities added by globalization.

For each domain Ui, define the local descent defect relation

Di($\mathcal{E}$) =
$C\mathcal{U}$,i($\mathcal{E}$) \ Ei.

A pair (x, y) belongs to Di($\mathcal{E}$) when

x, y ∈ Ui,

x and y are globally equivalent under $G\mathcal{U}(\mathcal{E})$,

but

x and y are locally distinct under Ei.

Faithful descent holds exactly when

Di($\mathcal{E}$) = ∅

for every i.

The defect relation provides a detailed certificate of incoherence.

A scalar penalty can be constructed from it.

For finite domains, a simple pair-count defect is

Lpair($\mathcal{E}$) =
Σi wi |Di($\mathcal{E}$)|,

where wi ≥ 0 weights the authority or importance of domain Ui.

Because ordered pairs may count both (x, y) and (y, x), one may instead count unordered pairs or normalize by domain size:

Lnorm($\mathcal{E}$) =
Σi wi |Di($\mathcal{E}$)| / |Ui|².

Pair count measures the number of local identity judgments reversed by global closure.

It can be sensitive to class size.

Merging two large local classes creates many defect pairs.

This may be appropriate because a large fusion destroys many distinctions.

A class-level defect may be more interpretable.

Let C be a global class of $G\mathcal{U}(\mathcal{E})$.

Let

mi(C) = number of Ei-classes intersecting C.

Faithful descent requires

mi(C) ≤ 1

for every i and C.

Define the collision penalty

Lcoll($\mathcal{E}$) =
Σi ΣC∈X/$G\mathcal{U}(\mathcal{E})$ wi max(0, mi(C) − 1).

This counts the number of excess local classes absorbed into global components.

A local object-loss measure is

Li($\mathcal{E}$) =
|Ui/Ei| − |Ui/$C\mathcal{U}$,i($\mathcal{E}$)|.

This is the number of local quotient objects lost after closure.

The total object-loss penalty is

Lobj($\mathcal{E}$) =
Σi wi Li($\mathcal{E}$).

Different penalties emphasize different consequences.

Pair-count loss emphasizes the volume of corrupted identity judgments.

Collision loss emphasizes the number of local classes fused.

Object loss emphasizes the reduction in local ontology size.

A witness-sensitive penalty can incorporate the length or reliability of the paths causing fusion.

For every defect pair (x, y), let

ℓ(x, y)

be the length of a shortest local identity zig-zag connecting them.

A penalty may assign greater weight to short paths because they indicate direct structural conflict, or greater weight to long paths because they reveal extensive identity propagation.

If identity edges carry confidence or cost, the witness penalty may depend on the weakest link, total path cost, or number of independent paths.

Thus a general descent penalty can be written

Ldes($\mathcal{E}$) =
Aggregate {
severity(x, y, path data) :
(x, y) ∈ ⋃i Di($\mathcal{E}$)
}.

The scalar loss should not replace the defect certificate.

Two families can have the same total penalty while failing in very different ways.

A rigorous optimization system should report both:

the numerical descent cost,

and the domains, local classes, and identity paths responsible for it.

5.5 Hard Versus Penalized Descent Constraints

There are two principal ways to incorporate descent into variational objecthood.

The first is a hard constraint:

C$\mathcal{U}$($\mathcal{E}$) = $\mathcal{E}$.

The optimization is restricted to closure-coherent families:

$\mathcal{E}$★ ∈ arg min {Jlocal($\mathcal{E}$) : $\mathcal{E}$ ∈ Fix(C$\mathcal{U}$)}.

No local distinction may be destroyed by globalization.

This formulation is appropriate when the local policies encode authoritative identity judgments.

Examples include:

certified disequalities in mathematics,

simultaneously observed distinct physical objects,

legally separate persons or entities,

safety-critical distinctions,

or experimentally established causal differences.

A single violated distinction makes the family infeasible.

The second approach uses a penalty:

$$J_{total}(\mathcal{E}) = J_{local}(\mathcal{E}) + \mu L_{des}(\mathcal{E}),$$

where $\mu \geq 0$ controls the cost of descent failure.

This formulation permits globally induced local fusion.

It is appropriate when local policies are uncertain, noisy, learned from incomplete data, or open to revision.

For small μ, the optimization may accept substantial local fusion to gain compression or predictive performance.

For large μ, the result approaches hard faithful descent.

The limit should be interpreted cautiously in discrete systems. A sufficiently large finite penalty may enforce the constraint when all other objective terms are bounded, but the required value depends on the problem.

Hard and penalized descent encode different ontological attitudes.

Hard descent gives local distinction priority.

The global object system must preserve every local classification.

Penalized descent allows global evidence to revise local ontology.

A local distinction can be sacrificed if the broader system gains enough.

Neither attitude is universally correct.

A local observer may possess decisive high-resolution information unavailable elsewhere.

In that case, global closure should not overrule it.

Alternatively, a local classifier may be noisy or over-refined. A global identity chain supported by many independent observers may justify merging its classes.

The weights w_i and μ can represent authority, reliability, or evidential strength.

However, scalar weighting can conceal substantive conflicts.

A highly weighted majority of weak observers should not necessarily override one observer possessing a direct causal test.

The theory may therefore require typed or lexicographic constraints.

Some distinctions are hard.

Others are soft.

Some identity paths may compose.

Others may not.

The present penalty formulation is the simplest approximate model.

A third approach introduces an explicit correction variable.

Let $\mathcal{E}$ be the observed local family and let $\mathcal{F}$ be a revised coherent family.

Solve

minimize

$$\mathrm{dlocal}(\mathcal{E}, \mathcal{F}) + \lambda K(\mathcal{F}) + D\pi(\mathcal{F})$$

subject to

$$C\mathcal{U}(\mathcal{F}) = \mathcal{F}.$$

Here dlocal measures the cost of revising the original local policies.

This separates observed identity judgments from the optimized coherent ontology.

The closure $C\mathcal{U}(\mathcal{E})$ is one possible revision, but it permits only coarsening.

A more general repair may both add and remove local identities.

The optimization can decide whether to:

accept globally forced fusion,

break one or more identity bridges,

refine a noisy local class,

or reorganize the policy family.

This repair formulation is more flexible than simply penalizing the closure defect.

It treats descent failure as evidence that the supplied policies themselves require revision.

5.6 Local Object Preservation

Faithful descent can be expressed directly at the quotient-object level.

For every domain Ui, define the canonical map

$\phi i: Ui/Ei \to X/G\mathcal{U}(\mathcal{E})$

by

$\phi i([x]Ei) = [x]G\mathcal{U}(\mathcal{E})$.

The map is always well defined because Ei is contained in the generated global relation.

Faithful descent holds exactly when every ϕi is injective.

Thus local object preservation is a no-collision condition.

Distinct local quotient objects from one observer must remain distinct globally.

Local objects from different observers may map to the same global object. That is the intended purpose of gluing.

The prohibited event is the merger of two objects already distinguished within one local domain.

This formulation suggests several preservation constraints.

A complete local preservation constraint requires injectivity of ϕi for every i.

A partial preservation constraint may designate only selected local classes as protected.

Let Pi be a set of ordered or unordered pairs of Ei-classes that must remain globally separate.

Then require

$\phi i(A) \neq \phi i(B)$

for every $(A, B) \in Pi$.

This permits some local coarsening while protecting high-value distinctions.

A weighted preservation loss may assign importance to local classes or class pairs.

If A and B merge globally, incur cost

w_i(A, B).

The total preservation penalty is

Lpres($\mathcal{E}$) =
Σi ΣA≠B in Ui/Ei
w_i(A, B) · 1[ϕi(A) = ϕi(B)].

This class-level formulation is often more efficient than counting state pairs.

It also corresponds more directly to the ontology of the observer.

A laboratory may care about preserving two biological types, not every pair of individual specimens.

A sensor may care about keeping two simultaneous tracks separate.

A classifier may care about protecting selected high-risk categories.

Local object preservation can also be dynamic.

Suppose each observer has an induced transformation on Ui/Ei.

A global object correspondence should preserve not only class injectivity but also the local transformation structure.

If two local objects merge globally, their future quotient behavior should remain compatible.

One may therefore require the local-to-global maps ϕi to be morphisms of dynamical systems, algebras, categories, or probabilistic models.

The strict set-level injection is only the first preservation layer.

A richer descent constraint would protect:

local object identity,

local operations,

local transition structure,

local causal response,

and local explanatory models.

This leads naturally toward categorical and stack-like variational descent.

5.7 Distributed Optimization over Observer Networks

A distributed objecthood problem contains both local and global objectives.

Each observer i may have its own distortion functional

Di(Ei)

and complexity functional

Ki(Ei).

The generated global relation may have its own distortion and complexity:

Dglobal(G$\mathcal{U}$($\mathcal{E}$)),

Kglobal(G$\mathcal{U}$($\mathcal{E}$)).

A general distributed objective is

Jdist($\mathcal{E}$) =
Σi αi[Di(Ei) + λiKi(Ei)]
+
βDglobal(G$\mathcal{U}$($\mathcal{E}$))
+
γKglobal(G$\mathcal{U}$($\mathcal{E}$))
+
μLdes($\mathcal{E}$).

The coefficients αi weight the local tasks.

The parameters λi control local compression.

The coefficients β and γ weight global performance.

The coefficient μ penalizes descent failure.

This objective captures several competing pressures.

Local observers prefer policies suited to their own data and tasks.

The global system prefers a coherent integrated ontology.

Local compression may conflict with global fidelity.

Global compression may destroy local objects.

A policy beneficial to one observer may create transitive consequences elsewhere.

Optimization must therefore be coordinated.

One architecture is local-first optimization.

Each observer independently chooses

$E_i^\star \in \arg\min [D_i(E_i) + \lambda_i K_i(E_i)]$.

The resulting family is then globalized and tested.

This architecture is decentralized but vulnerable to descent obstruction.

Independent local optima need not form a globally coherent family.

A second architecture is global-first optimization.

Choose one global relation E on X and define every local policy by restriction:

$E_i = E|U_i$.

Descent coherence is automatic.

The optimization becomes

$$E^\star \in \arg\min \{ D_{global}(E) + \lambda K_{global}(E) + \Sigma_i \alpha_i [D_i(E|U_i) + \lambda_i K_i(E|U_i)] \}.$$

This avoids gluing obstruction because the local policies are induced by one global ontology.

It may be computationally harder and may grant insufficient autonomy to local observers.

A third architecture is joint consensus optimization.

Treat the global policy E and the local policies Ei as separate variables:

minimize

$D_{global}(E) + \lambda K_{global}(E)$

-

$\Sigma_i \alpha_i [D_i(E_i) + \lambda_i K_i(E_i)]$

-

μΣi d(Ei, E|Ui).

The function d measures disagreement between each local policy and the restriction of the global policy.

As μ increases, local policies approach exact global consistency.

This resembles consensus optimization in distributed computation.

It also permits observer-specific noise, uncertainty, and task differences.

A fourth architecture optimizes generators.

Rather than selecting complete equivalence relations directly, observers propose weighted identity edges. The global system selects a subset, computes equivalence closure, and evaluates the resulting quotient.

This approach is natural for entity resolution, multisensor tracking, and knowledge-graph integration.

The optimization must account for the nonlinear effect of closure.

One accepted bridge can merge large components and trigger substantial local loss.

The marginal cost of an edge therefore depends on the entire current identity graph.

Distributed optimization also raises questions of authority and information access.

An observer may not reveal its complete local state space.

It may transmit only quotient classes, summaries, or protected constraints.

The global system must form objects without centralized raw data.

The local-object graph supports such computation.

Each node represents a local quotient object.

Edges represent proposed cross-observer identity.

A global component is feasible when it contains no prohibited local collision.

Optimization can then operate over correspondence edges rather than individual states.

This produces a practical route toward federated object discovery.

5.8 Closure-Stable Feasible Regions

The exact distributed feasible region is

Ades =
Fix(C$\mathcal{U}$).

If transformation invariance, task preservation, and protected distinctions are also required, the complete feasible region becomes

Aπ =
Fix(C$\mathcal{U}$)
∩
Inv(G)
∩
Pres(π)
∩
Sep(P).

The variational problem is

$\mathcal{E}$★ ∈ arg min {Dπ($\mathcal{E}$) + λK($\mathcal{E}$) : $\mathcal{E}$ ∈ Aπ}.

The fixed-point family Fix(C$\mathcal{U}$) is structurally well behaved.

It is closed under arbitrary meets.

If several coherent distributed ontologies are available, their componentwise common refinement remains coherent.

This makes refinement a safe operation with respect to closure stability when applied through meets of complete coherent families.

Joins require renewed closure.

The componentwise join of two fixed-point families may not itself be fixed.

The coherent join is

$\mathcal{E}$ ∨coh $\mathcal{F}$ =
C$\mathcal{U}$($\mathcal{E}$ ∨ $\mathcal{F}$),

where the join inside the parentheses is taken in the ambient product lattice.

Thus the closure operator repairs the result of merging distributed policies.

The fixed-point family is itself a complete lattice under these operations.

This provides an exact search space for globally coherent object systems.

One can move downward through coherent refinements.

One can move upward through closure-completed coarsenings.

Every arbitrary local family has a least coherent coarsening $C\mathcal{U}(\mathcal{E})$.

This supports a projection-based optimization strategy.

Given a proposed update $\mathcal{E}'$, replace it with

$C\mathcal{U}(\mathcal{E}')$.

The update is returned to the coherent region.

However, the projection may erase local distinctions.

A gradient-like or heuristic search using this operation must include the resulting descent cost in its objective.

Otherwise, every incoherent proposal is silently repaired through coarsening, potentially producing false fusion.

An alternative is a repair-by-refinement projection.

Instead of accepting all identity declarations and adding their consequences, remove selected generators until faithful descent is restored.

This projection is not unique.

It requires an optimization criterion, such as minimum removed weight or maximum retained evidence.

Thus two directions of repair exist.

Closure repair moves upward:

add globally forced identities.

Separation repair moves downward:

remove identities responsible for forbidden fusion.

The first preserves all positive identity declarations.

The second preserves selected distinctions.

Variational objecthood must decide which information has priority.

The closure-stable feasible region can also be parameterized by global relations.

Every $E \in Eq(X)$ generates a coherent local family

$R_{\mathcal{U}}(E) = (E|U_i)_i$.

Not every distinct global E yields a distinct local family. Optional identities invisible to the cover may disappear under restriction.

For optimization focused only on local behavior, such global policies may be observationally equivalent.

For global prediction or dynamics, the distinction may matter.

Thus one may optimize either:

over coherent local families,

over minimally generated global relations,

or over all global relations consistent with the local family.

These are different search spaces.

The minimal generated relation contains only identities forced by local data.

Larger global relations add optional identities.

A least-complexity objective may favor some of those additional mergers if they preserve local injectivity and task fidelity.

The realization category from the descent theory therefore becomes a secondary variational space.

For a fixed coherent local family $\mathcal{E}$, one may solve

$$E^{\bigstar} \in \arg\min \{ D_{\pi}(E) + \lambda K(E) : E|U_i = E_i \text{ for every } i \}.$$

The minimal globalization $G_{\mathcal{U}}(\mathcal{E})$ is the finest feasible relation.

A coarser faithful realization may achieve more compression without damaging local policies.

Global identity selection therefore continues even after local descent is certified.

The principal conclusions of this section are:

Distributed identity policies form a product lattice of local equivalence relations.

Globalization followed by restriction defines a closure operator $C_{\mathcal{U}}$.

Closure-coherent families are exactly the fixed points of $C\mathcal{U}$.

The fixed-point equation is equivalent to faithful descent.

Descent defects measure the identities added locally by global closure.

These defects can be treated as hard violations or weighted penalties.

Local object preservation is equivalent to injectivity of local-to-global quotient maps.

Distributed optimization may be local-first, global-first, consensus-based, or generator-based.

The exact coherent feasible region is a complete lattice of closure fixed points.

Every local family has a canonical least coherent coarsening, but this coarsening may destroy locally important distinctions.

The next section adds transformation invariance to the feasible region. It develops exact and approximate invariance, congruence-like identity policies, induced quotient dynamics, and the lattice of dynamically admissible object systems.

6. Transformation Invariance as an Admissibility Constraint

A quotient object must survive the transformations relevant to the theory in which it is used.

If two states are declared equivalent, their future images, transformed forms, or operational consequences must remain equivalent. Otherwise the quotient cannot assign one unambiguous transformed state to their common object class.

This requirement is transformation invariance.

Let E be an equivalence relation on a state space X.

Let g: X → X be an admissible transformation.

The policy E is invariant under g when

x E y implies g(x) E g(y).

When this condition holds, g induces a unique quotient transformation

$\bar{g}$: X/E → X/E

defined by

$\bar{g}$([x]E) = [g(x)]E.

The induced map is well defined because the image class does not depend on which representative x is chosen.

If the invariance condition fails, there exist states x and y in one E-class whose transformed images belong to different E-classes. The proposed object then has more than one possible transformed image. The quotient dynamics are representative-dependent and therefore not lawful at the object level.

Transformation invariance is stronger than visual similarity, statistical clustering, or static agreement.

Two states can look similar while evolving differently.

Two records can be classified together while responding differently to an update.

Two physical configurations can share selected observables while diverging under time evolution.

Two biological states can occupy one descriptive class while responding differently to intervention.

A quotient intended to support transformation must preserve the distinctions needed for that transformation.

The family of admissible transformations therefore constrains the possible identity policies.

A relation valid under one transformation regime may fail under another.

A quotient appropriate for short-term prediction may be unstable over long times.

A policy suitable for passive observation may fail under intervention.

A classification stable under rotations may fail under changes of scale.

Objecthood is therefore transformation-relative.

This section defines admissible transformation families, invariant equivalence relations, and congruence-like identity policies. It develops induced quotient dynamics, exact and approximate invariance, defect measures, and the order structure of dynamically admissible policies.

6.1 Admissible Transformation Families

Let X be a state space.

A transformation is a function

$g: X \to X$.

The function may represent:

time evolution,

an action taken by an agent,

a physical symmetry,

a computational update,

a measurement procedure,

a perturbation,

a change of representation,

or another operation under which object identity is expected to persist.

Not every mathematically definable function should be treated as admissible.

The relevant family must be selected according to the task policy π.

Let G be a collection of admissible transformations on X.

Several algebraic structures are possible.

G may be a set of independent transformations.

It may be a monoid containing the identity transformation and closed under composition.

It may be a group when every transformation is invertible.

It may be a semigroup when closure under composition is present but an identity element is not required.

It may be a category or groupoid when transformations have different source and target domains.

It may be a stochastic family rather than a collection of deterministic functions.

The monoid case is the most direct foundation.

Assume G contains idX and that for g, h ∈ G,

g ∘ h ∈ G.

A policy E is G-invariant when it is preserved by every g ∈ G:

x E y implies g(x) E g(y)

for all x, y ∈ X and g ∈ G.

The set of all such policies is denoted

Inv(G).

The selection of G determines what the objects must survive.

If G contains only the identity transformation, every equivalence relation is invariant.

If G contains many transformations capable of revealing hidden differences, fewer policies remain admissible.

Adding transformations to G can only shrink Inv(G).

If

$G \subseteq H$,

then

$\mathrm{Inv}(H) \subseteq \mathrm{Inv}(G)$.

A policy stable under the larger family H is automatically stable under the smaller family G.

The converse need not hold.

Thus transformation-rich tasks generally require finer identity policies.

The family G may be generated by a smaller set S.

Suppose every $g \in G$ can be written as a finite composition of generators in S.

If E is preserved by every generator $s \in S$, then E is preserved by every $g \in G$.

This follows because preservation is closed under composition.

Therefore invariance can be tested on generators rather than on every transformation in the monoid.

This can substantially reduce computational cost.

The admissible family must be distinguished from transformations that merely happen to occur in the data.

A rare nuisance transformation may be irrelevant to the task.

A possible intervention may be critical even if it has not yet been observed.

The family G is therefore partly empirical and partly policy-driven.

For image identity, G may contain rotations, translations, changes in illumination, or deformations regarded as preserving the depicted object.

For a dynamical system, G may contain time-evolution maps.

For control, G may contain all permitted actions.

For mathematics, G may contain structure-preserving maps or rewrite operations.

For scientific explanation, G may contain transformations under which the proposed object laws should remain valid.

A policy invariant under a carefully selected G expresses a strong claim:

Every admissible operation respects the identity classes.

The quotient objects are stable units under the transformation regime.

6.2 Invariant Equivalence Relations

For one transformation $g: X \to X$, define

$Inv(g) =$
$\{E \in Eq(X) : x \, E \, y \text{ implies } g(x) \, E \, g(y)\}.$

For a family G,

$Inv(G) = \bigcap_{g \in G} Inv(g).$

Thus G-invariant relations are exactly those preserved by every admissible transformation.

Several elementary facts follow.

The equality relation ΔX belongs to Inv(G) for every G.

If $x = y$, then $g(x) = g(y)$.

The universal relation ∇X also belongs to Inv(G).

Every pair of transformed states remains universally equivalent.

These two extreme policies are always dynamically admissible.

Their presence again shows that invariance alone cannot select useful objects.

The meet of invariant equivalence relations is invariant.

Let $\{E\alpha\}$ be a family in Inv(G), and let

$E = \bigcap_{\alpha} E\alpha.$

If x E y, then x Eα y for every α.

Because every Eα is invariant,

g(x) Eα g(y)

for every α.

Therefore

g(x) E g(y).

Thus Inv(G) is closed under arbitrary meets.

The join of invariant equivalence relations is also invariant.

Let

E = EqCl(⋃α Eα).

If x E y, there is a finite chain

x = x0, x1, …, xn = y

in which each consecutive pair is related by some Eα.

Applying g gives

g(x0), g(x1), …, g(xn).

Each consecutive transformed pair remains related by the same invariant Eα.

Therefore g(x) E g(y).

Thus Inv(G) is closed under arbitrary joins.

Consequently, Inv(G) is a complete sublattice of Eq(X).

This result is important.

The dynamically admissible policies are not an arbitrary feasible set. They retain the full meet-and-join structure of the policy lattice.

Their meet is ordinary intersection.

Their join is ordinary equivalence closure of the union.

No additional dynamical repair is required because the generators are already invariant.

The finest invariant policy is equality.

The coarsest invariant policy is universal equivalence.

Task and distortion constraints are therefore needed to identify useful intermediate points.

A transformation can induce a closure operation on arbitrary relations.

Given a relation R, one may seek the smallest G-invariant equivalence relation containing R.

Define

InvClG(R)

as the intersection of all G-invariant equivalence relations containing R.

Because Inv(G) is closed under intersections, this relation exists.

It is the invariant equivalence closure generated by R.

Operationally, it can be constructed by adding:

the equivalence consequences of R,

the transformed pairs g(x), g(y) whenever x R y,

and all consequences generated by repeated transformation and equivalence closure.

The resulting relation is the least identity policy accepting the initial declarations and remaining stable under G.

This construction parallels distributed globalization.

Local identity closure adds consequences of transitive paths.

Invariant closure adds consequences of admissible transformation.

A policy satisfying both descent and invariance must be closed under both operations.

The interaction of these closures will become important in the variational principle.

6.3 Congruence-Like Identity Policies

In abstract algebra, a congruence is an equivalence relation compatible with the operations of an algebraic structure.

If X carries a binary operation $\star$, a congruence E satisfies

x E x′ and y E y′ imply x $\star$ y E x′ $\star$ y′.

The quotient X/E then inherits a well-defined operation:

[x]E $\bar{\star}$ [y]E = [x $\star$ y]E.

Transformation invariance generalizes this idea.

Each admissible transformation can be viewed as an operation that must descend to the quotient.

If G consists of unary operations, G-invariant equivalence relations are precisely congruences for the unary algebra generated by G.

If the state space carries operations of several arities, admissibility requires compatibility with all of them.

Let ω be an n-ary operation

$\omega: X^n \to X$.

An equivalence relation E is compatible with ω when

x1 E y1, …, xn E yn

imply

ω(x1, …, xn) E ω(y1, …, yn).

The operation then descends to X/E:

$\bar{\omega}([x1]E, \ldots, [xn]E)$

[ω(x1, …, xn)]E.

An identity policy compatible with every operation in a signature Ω is an Ω-congruence.

This broader formulation is necessary for object systems supporting combination, composition, interaction, or algebraic computation.

Examples include:

addition and multiplication in quotient rings,

composition in quotient categories,

logical operations in quotient algebras,

combination of physical subsystems,

and program constructors in semantics.

The word congruence emphasizes that object identity must be respected compositionally.

If equivalent parts can be substituted into a larger construction and produce inequivalent results, the identity policy is not admissible for that construction.

This gives a powerful interpretation of objecthood:

An object class is a lawful substitution class for every operation admitted by the theory.

Transformation invariance is the unary case of this substitution principle.

For a task involving both dynamics and composition, the feasible set should be the lattice of congruences for the full operation family.

Write

$\mathrm{Con}\Omega(X)$

for this lattice.

Then the variational problem becomes

$E\bigstar \in \arg\min \{D\pi(E) + \lambda K(E) : E \in \mathrm{Con}\Omega(X) \text{ and other constraints hold}\}.$

Congruence lattices are central in universal algebra.

The present theory interprets them as spaces of operation-compatible object policies.

This connection places variational objecthood within a broad algebraic tradition while adding the selection problem absent from ordinary congruence theory.

Classical algebra asks which congruences exist and what quotients they produce.

Variational objecthood asks which congruence produces the best effective object system for a selected purpose.

6.4 Representative-Independent Quotient Dynamics

Suppose $E \in \mathrm{Inv}(G)$.

For each $g \in G$, define

$\bar{g}: X/E \to X/E$

by

$\bar{g}([x]E) = [g(x)]E.$

This map is well defined.

If $[x]E = [y]E$, then $x \mathrel{E} y$.

Invariance gives

g(x) E g(y).

Therefore

[g(x)]E = [g(y)]E.

The quotient map qE satisfies

qE ∘ g = ḡ ∘ qE.

This is the equivariance equation.

It states that transforming a detailed state and then quotienting gives the same result as quotienting first and transforming the effective object.

The square formed by X, X/E, g, ḡ, and qE commutes.

If G is a monoid, the induced maps form a monoid action on X/E.

For the identity transformation,

id̄([x]E) = [x]E.

For g, h ∈ G,

(g ∘ h̄)([x]E)

[g(h(x))]E

and

ḡ(h̄([x]E))

ḡ([h(x)]E)

[g(h(x))]E.

Therefore

(g ∘ h̄) = ḡ ∘ h̄.

The quotient preserves the algebraic organization of the transformation family.

This is more than a technical convenience.

It means that reasoning at the object level is closed.

One need not return to detailed representatives after every transformation.

The quotient possesses its own lawful dynamics.

A useful effective theory requires this autonomy.

Suppose E is not invariant.

Then there exist x E y such that

g(x) and g(y)

belong to different E-classes.

The expression

$\bar{g}([x]E)$

has at least two possible values depending on whether x or y is chosen as representative.

No deterministic quotient map exists.

Several responses are possible.

Refine E so that x and y become distinct objects.

Restrict the transformation family so that g is no longer required.

Replace deterministic quotient dynamics with a set-valued or probabilistic transition.

Enrich the quotient object with hidden state or memory.

Accept an approximation and measure the resulting distortion.

These responses correspond to different modeling choices.

Refinement preserves deterministic autonomy at the cost of complexity.

Relaxation preserves compression at the cost of uncertainty or history dependence.

The variational theory can compare these alternatives when the corresponding costs are represented.

Representative independence also applies to observables and interventions.

A quotient-level prediction

$\bar{p}: X/E \to Y$

exists exactly when p is constant on E-classes.

A quotient-level action response exists exactly when equivalent states respond equivalently.

A quotient-level operation exists exactly when E is a congruence.

Thus representative independence is the unifying criterion for lawful abstraction.

The effective object should support the intended functions without requiring knowledge of which detailed state lies inside it.

6.5 Exact Invariance

Exact invariance requires

x E y implies g(x) E g(y)

without exception.

This condition is appropriate when quotient operations must be mathematically well defined.

Exact quotient algebras, formal state minimization, symbolic computation, and deterministic abstract machines often require this standard.

For finite systems, exact invariance can be tested directly.

Let the E-classes be

C1, …, Ck.

For every g ∈ G and every class Cj, compute the image set

g(Cj) = {g(x) : x ∈ Cj}.

The policy is invariant under g exactly when g(Cj) lies entirely within one E-class for every j.

Equivalently, the set of quotient labels

{[g(x)]E : x ∈ Cj}

must have size one.

If G has a finite generating set S, test only the generators.

A class-level defect certificate consists of:

a class C,

a transformation g,

and states x, y ∈ C

such that g(x) and g(y) fall into different quotient classes.

This certificate identifies precisely why the quotient action fails.

Exact invariance can also be expressed through kernel refinement.

For a fixed g and policy E, define the pullback relation

$g^{-1}(E)$

by

$x\ g^{-1}(E)\ y$

exactly when

$g(x)\ E\ g(y)$.

Then E is invariant under g exactly when

$E \subseteq g^{-1}(E)$.

For a family G,

$E \in Inv(G)$

exactly when

$E \subseteq \bigcap_{g \in G} g^{-1}(E)$.

This formulation connects invariance with fixed-point methods.

Define an operator

$TG(E) = E \cap \bigcap_{g \in G} g^{-1}(E)$.

The result removes pairs whose images are not E-equivalent.

Repeated application may refine E toward a G-invariant relation.

For finite X, iteration eventually stabilizes.

The greatest invariant relation contained in a supplied E can be obtained by such refinement procedures under suitable definitions.

This is a separation-preserving repair.

The dual repair seeks the least invariant relation containing E through InvClG(E).

Thus invariance failure can be repaired in two directions.

Invariant coarsening adds identity consequences until transformed images remain together.

Invariant refinement removes unstable identity declarations until each class evolves coherently.

Coarsening may increase distortion by merging additional states.

Refinement increases complexity.

The variational objective decides which repair is preferable.

Exact invariance may be required only for a selected operation family.

A quotient can be exact under one set of transformations and approximate under another.

The feasible set should therefore distinguish mandatory operations from optional or noisy ones.

Let Ghard be the hard transformation family and Gsoft the soft family.

Require

$E \in \mathrm{Inv}(G_{\mathrm{hard}})$,

and penalize failures under Gsoft.

This mixed formulation is often more realistic than treating every transformation identically.

6.6 Approximate Invariance

Exact invariance is frequently too strict for empirical and continuous systems.

Measurements are noisy.

Learned transitions are uncertain.

Objects deform.

Prediction errors accumulate.

States within one useful object class may not evolve identically, yet their quotient-level behavior may remain sufficiently similar.

Approximate invariance replaces binary preservation with a graded defect.

Let $q_E: X \to X/E$ be the quotient map.

Suppose the quotient object space carries a distance d_E or a task-relevant loss between classes.

For one transformation g, define the class image spread

$$\delta_g(C) = \sup \{d_E(q_E(g(x)), q_E(g(y))) : x, y \in C\}.$$

The worst-case invariance defect is

$S_{inv,max}(E) =$
$\sup \{\delta_g(C) : C \in X/E \text{ and } g \in G\}.$

If d_E is the binary distance

$d_E(A, B) =$
0 when $A = B$,

1 when $A \neq B$,

then $S_{inv,max}(E) = 0$ exactly when E is exactly invariant.

For metric or probabilistic quotient structures, d_E can measure the severity of representative disagreement.

An expected invariance defect is

$S_{inv,avg}(E) =$
$E[d_E(q_E(g(x)), q_E(g(y))) \mid x \, E \, y]$,

with averaging over transformations, states, and equivalent pairs according to a selected distribution.

A decoder-based formulation may choose the best quotient transformation.

For each g, seek a map

$\hat{g}: X/E \to X/E$

minimizing

$E[L(q_E(g(x)), \hat{g}(q_E(x)))]$.

The minimal loss is the approximation error of representing g at the quotient level.

Define

$S_g(E) =$
$\inf$ over $\hat{g}$
$E[L(q_E(g(x)), \hat{g}(q_E(x)))]$.

Then

$S_{inv}(E) =$
Aggregate $\{S_g(E) : g \in G\}$.

This formulation does not require a predetermined quotient transformation.

It asks how well the detailed transformation can be approximated by any object-level rule.

Zero error recovers exact factorization:

$qE \circ g = \hat{g} \circ qE$.

Positive error measures failure of autonomous quotient dynamics.

For stochastic systems, the quotient-level decoder is a transition kernel rather than a function.

Equivalent states may induce slightly different distributions over quotient objects.

A divergence between those distributions provides a stochastic invariance defect.

Approximate invariance permits useful coarse-graining.

A macroscopic variable may not be perfectly Markovian but may predict sufficiently well for the intended time scale.

An image class may deform under transformation but remain recognizable with high probability.

A biological type may exhibit internal variation while preserving a common response distribution.

The acceptable level of defect depends on π.

Approximate invariance also supports a penalized objecthood functional:

$$J(E) = D\pi(E) + \lambda K(E) + \nu Sinv(E),$$

where ν controls the cost of dynamical instability.

If exact invariance is mandatory, restrict E to Inv(G) instead.

If approximate autonomy is acceptable, the penalty permits a smoother tradeoff.

One must avoid double counting.

Predictive distortion and invariance defect may measure related failures.

$D\pi$ should quantify task loss, while Sinv should quantify failure of structural closure under transformation.

Their definitions should be coordinated.

6.7 Invariance Defects

An invariance defect identifies how a proposed object fails under transformation.

For a deterministic transformation g and equivalence relation E, define the defect set

Defg(E) =
{(x, y) ∈ E : g(x) not E g(y)}.

The policy is invariant under g exactly when

Defg(E) = ∅.

For a family G,

DefG(E) =
{(g, x, y) :
g ∈ G,
x E y,
and g(x) not E g(y)}.

This set contains complete certificates of representative dependence.

Several scalar defect measures can be derived.

Pair-count defect:

Spair(E) =
Σg∈G wg |Defg(E)|.

Normalized pair defect:

Snorm(E) =
Σg∈G wg |Defg(E)| / |E|,

when E is finite and nonempty.

Class-splitting defect:

For each class C and transformation g, let

mg(C) =
number of E-classes intersecting g(C).

Exact invariance requires

$mg(C) = 1$.

Define

$$S_{split}(E) = \sum_{g \in G} \sum_{C \in X/E} w_g \max(0, m_g(C) - 1).$$

This counts how many distinct quotient images one object acquires.

A class may split into two, three, or more images depending on the representative.

The measure therefore has a direct object-level interpretation.

A state-weighted defect can assign greater importance to common or critical states.

A transformation-weighted defect can emphasize operations central to the task.

A temporal defect can weight early or late divergence differently.

A severity-weighted defect can measure how far apart the image classes are rather than merely whether they differ.

A complete diagnostic should identify:

the source quotient class,

the transformation causing failure,

the representatives producing different images,

the resulting quotient classes,

and the task consequences of the disagreement.

This information guides repair.

Suppose a class C splits under g into image classes D_1 and D_2.

One repair is to split C according to which image class each state reaches.

Define

$$C_1 = \{x \in C : g(x) \in D_1\},$$

$$C_2 = \{x \in C : g(x) \in D_2\}.$$

Refining C in this way removes the one-step defect for g.

Further iterations may be needed because the new classes can still split under other transformations or later compositions.

This process resembles partition refinement in automata minimization and bisimulation computation.

Another repair is to merge D1 and D2.

This preserves C as one object under g but coarsens the future ontology.

The merger may propagate through the system and increase task distortion.

A third repair is to replace the deterministic quotient map with a probabilistic rule assigning a distribution over D1 and D2.

This retains the coarse object but acknowledges internal uncertainty.

The repair choice reveals the interpretation of the original class.

If internal representatives truly have different lawful futures, refinement may be required.

If D1 and D2 are distinctions irrelevant to the task, coarsening may be justified.

If the difference is stochastic or hidden, probabilistic dynamics may be appropriate.

Invariance defects therefore do more than reject policies.

They identify the precise tension among compression, deterministic autonomy, and retained detail.

6.8 The Lattice of Dynamically Admissible Policies

The family Inv(G) of G-invariant equivalence relations forms a complete lattice under the refinement order.

This lattice is the natural dynamical feasible space for variational objecthood.

Its bottom element is equality.

Its top element is universal equivalence.

Its meet is intersection.

Its join is equivalence closure of union.

Because all generating relations are invariant, their join remains invariant.

The lattice can be narrower than Eq(X), sometimes dramatically so.

A transformation family can force identity policies to align with dynamical behavior.

For deterministic finite-state systems, invariant relations group states so that each class has one well-defined successor class.

For action systems, invariant relations require one well-defined successor class under every action.

For algebraic systems, congruence lattices require compatibility with every operation.

The structure of Inv(G) reflects the transformation system.

A transformation family that strongly separates states may leave few nontrivial invariant policies.

A highly symmetric system may possess many natural invariant quotients.

A dynamically admissible policy can be moved upward through invariant coarsening or downward through invariant refinement while remaining in Inv(G), provided lattice operations are used.

This supports constrained search.

One may begin at equality and search upward through invariant merges.

One may begin at universal equivalence and search downward through invariant splits.

One may project a candidate policy onto:

the least invariant coarsening containing it,

or the greatest invariant refinement contained in it, when such a refinement is constructed.

The choice again corresponds to whether positive identities or preserved distinctions receive priority.

Transformation invariance must also be combined with closure coherence.

Let Fix($C\mathcal{U}$) be the lattice of distributed closure fixed points.

The exact feasible region is

A = Fix($C\mathcal{U}$) ∩ Inv(G),

with suitable interpretation when E is a distributed family.

An intersection of two complete sublattices need not automatically preserve every join inherited from the ambient lattice without closure under both operations. However, it remains closed under meets when both families are meet-closed.

A joint closure operator can be constructed by alternating or composing descent closure and invariant closure until a common fixed point is reached.

Let Cdes represent distributed closure and Cinv represent invariant coarsening.

Starting from a policy E0, define

E1 = Cdes(Cinv(E0)),

E2 = Cdes(Cinv(E1)),

and continue.

For finite policy lattices, the ascending sequence stabilizes.

The limit is a common coherent coarsening under suitable monotonicity assumptions.

The resulting relation contains all identities forced by both distributed gluing and transformation stability.

It may be substantially coarser than the initial policy.

A separation-preserving approach would instead seek a common refinement satisfying both conditions.

The variational objective should compare the distortion and complexity of these alternatives.

The full constrained problem is

$$
\begin{aligned}
&E^{\bigstar} \in \arg\min \{ \\
&D\pi(E) + \lambda K(E) : \\
&E \in \mathrm{Fix}(C\mathcal{U}) \cap \mathrm{Inv}(G) \\
&\}.
\end{aligned}
$$

A penalized version is

$$
\begin{aligned}
&E^{\bigstar} \in \arg\min \{ \\
&D\pi(E) \\
&+ \\
&\lambda K(E) \\
&+ \\
&\mu L_{\mathrm{des}}(E) \\
&+ \\
&\nu S_{\mathrm{inv}}(E) \\
&:
\end{aligned}
$$

$E \in \mathrm{Eq}(X)$
}.

The hard form selects only perfectly coherent and invariant policies.

The penalized form permits approximate objecthood.

These equations give the variational theory its structural foundation.

Complexity rewards compression.

Distortion protects task-relevant differences.

Descent protects local coherence.

Invariance protects lawful transformation.

An optimized policy must balance all four.

The principal conclusions of this section are:

Admissible transformations determine which identities must persist.

Invariant equivalence relations are exactly those supporting representative-independent quotient maps.

Transformation families may form monoids, groups, categories, or stochastic systems.

Invariant relations generalize algebraic congruences.

Inv(G) is a complete lattice of dynamically admissible policies.

Exact invariance gives autonomous quotient dynamics.

Approximate invariance measures how well detailed transformations can be represented at the object level.

Defect certificates identify classes whose transformed image depends on the representative.

Invariance can be repaired by refinement, coarsening, probabilistic dynamics, or restriction of the transformation family.

The intersection of closure coherence and transformation invariance defines the structural feasible region of variational objecthood.

The next section states the complete variational objecthood principle. It combines distortion, complexity, descent, and invariance into one constrained optimization problem and interprets that problem as an algebraic least-action principle for the selection of effective objects.

7. The Variational Objecthood Principle

The preceding sections have identified the elements required for object selection.

The state space X supplies the detailed states.

The lattice Eq(X) supplies the candidate identity policies.

The distortion functional Dπ(E) measures the task-relevant damage caused by quotienting.

The complexity functional Kπ(E) measures the burden of representing and using the resulting object system.

The closure condition protects local-to-global coherence.

The transformation condition protects representative-independent dynamics.

These components can now be combined into one variational principle.

The central problem is not simply to find an equivalence relation.

It is to find an equivalence relation that produces the most economical effective ontology while preserving the observations, transformations, distinctions, and outcomes that matter.

Let Aπ be the family of identity policies admissible under task policy π.

A basic exact formulation is

$E^{\bigstar} \in \arg\min \{D\pi(E) + \lambda K\pi(E) : E \in A\pi\}$.

The feasible family may be defined by

$$A\pi = \mathrm{Fix}(C\mathcal{U}) \cap \mathrm{Inv}(G) \cap \mathrm{Pres}(H) \cap \mathrm{Sep}(P),$$

where:

$\mathrm{Fix}(C\mathcal{U})$ contains closure-coherent policies,

Inv(G) contains transformation-invariant policies,

Pres(H) contains policies preserving required observables,

and Sep(P) contains policies preserving protected distinctions.

The parameter $\lambda \geq 0$ expresses the cost of complexity relative to distortion.

A large λ favors stronger compression.

A small λ favors greater fidelity.

This equation is the variational objecthood principle.

It converts object formation from a supplied act of classification into a mathematical selection problem.

The resulting quotient objects are not chosen merely because their relation is coherent. They are selected because the relation achieves an optimal balance between lawful substitution and retained complexity.

Several formulations of the principle are possible.

One may impose coherence and invariance as exact constraints.

One may penalize violations.

One may minimize complexity under a fixed distortion tolerance.

One may maximize compression under preservation requirements.

One may seek the Pareto frontier of all nondominated identity policies.

These are not rival theories. They are different mathematical expressions of the same object-selection problem.

This section states the master optimization problem, develops constrained and penalized formulations, introduces multiplier methods, defines the full objecthood functional, characterizes least-complexity admissible quotients, and explains the analogy with least-action principles.

7.1 Statement of the Master Optimization Problem

Let X be a state space.

Let Eq(X) be the complete lattice of equivalence relations on X.

Let π specify the purpose for which objects are being formed.

The policy π may include:

required observables,

prediction targets,

intervention families,

time horizons,

error tolerances,

representation symmetries,

protected distinctions,

observer domains,

and admissible transformations.

For each $E \in Eq(X)$, define:

$D\pi(E)$ = task-relevant distortion,

$K\pi(E)$ = effective object-system complexity.

Let $A\pi \subseteq Eq(X)$ be the feasible family.

The master optimization problem is

$$E\bigstar \in \arg\min \{J\pi,\lambda(E) : E \in A\pi\},$$

where

$$J\pi,\lambda(E) = D\pi(E) + \lambda K\pi(E).$$

The solution $E\bigstar$ is an optimal identity policy.

Its quotient

$$X/E\bigstar$$

is the selected effective object space.

The equivalence classes of $E\bigstar$ are the objects produced by the variational principle.

The formulation separates three mathematical roles.

$Eq(X)$ defines what identity policies are possible.

$A\pi$ defines which policies are lawful.

$J\pi,\lambda$ ranks the lawful policies.

This distinction is essential.

A low-cost policy outside Aπ is not a valid solution.

A high-compression policy that violates transformation invariance does not support lawful quotient dynamics.

A low-distortion policy that violates descent coherence does not form a consistent distributed ontology.

A policy that merges a protected pair is inadmissible regardless of its numerical efficiency when the separation constraint is hard.

The feasible set may be written explicitly as

Aπ =
{E ∈ Eq(X) :
C$\mathcal{U}$(E) = E,
E ∈ Inv(G),
E ⊆ EH,
and P ∩ E = ∅},

where EH is the common kernel of all exactly preserved observables and P is a set of protected unequal pairs.

The notation must be adapted when C$\mathcal{U}$ acts on a family of local policies rather than one global relation. In the distributed case, the optimization variable is

$\mathcal{E}$ = (Ei)i∈I,

and the master problem becomes

$\mathcal{E}$★ ∈ arg min {Jπ,λ($\mathcal{E}$) : $\mathcal{E}$ ∈ Aπ,$\mathcal{U}$}.

The selected global relation is then

E★global = G$\mathcal{U}$($\mathcal{E}$★).

The local and global formulations answer different questions.

The global formulation searches directly for one equivalence relation on X.

The distributed formulation allows observers to retain local policy variables while requiring or encouraging coherent globalization.

The master principle is intentionally general.

Different applications define different Dπ, Kπ, and Aπ.

The mathematical architecture remains the same.

A policy is preferred when no other lawful policy achieves a lower total cost.

If the minimizer is unique, the task and constraints select one effective ontology.

If several minimizers exist, the problem admits several equally optimal object systems.

If the feasible set is empty, no exact object system satisfies all requirements.

If the infimum exists but is not attained in an infinite setting, one obtains approximating sequences rather than a selected policy.

These possibilities must be distinguished.

The variational principle does not guarantee uniqueness or even existence without assumptions.

Finite state spaces provide the simplest exact setting because Eq(X) is finite.

7.2 Constrained and Penalized Formulations

Hard constraints and penalties express different kinds of admissibility.

A hard-constrained formulation permits only policies satisfying every required structural law.

For example,

$E^{\bigstar} \in \arg\min \{D\pi(E) + \lambda K\pi(E) :$
$E \in \mathrm{Fix}(C\mathcal{U}) \cap \mathrm{Inv}(G)\}.$

This formulation rejects all descent and invariance violations.

It is appropriate when the quotient must support exact representative-independent dynamics and preserve every authoritative local distinction.

A penalized formulation allows violations but assigns them cost.

Define:

Ldes(E) = descent defect,

Sinv(E) = invariance defect,

Lsep(E) = protected-separation defect,

Lobs(E) = exact-observable preservation defect.

Then define

$$J_{total}(E) = D_{\pi}(E) + \lambda K_{\pi}(E) + \mu L_{des}(E) + \nu S_{inv}(E) + \rho L_{sep}(E) + \sigma L_{obs}(E).$$

The optimal policy is

$$E^{\bigstar} \in \arg\min \{J_{total}(E) : E \in Eq(X)\}.$$

The coefficients μ, ν, ρ, and σ control the severity of the corresponding violations.

This formulation is useful when:

the observations are noisy,

the local policies are provisional,

the transformation model is approximate,

the protected distinctions carry confidence levels,

or no exactly feasible policy exists.

Hard constraints can be represented within the objective by extended-valued penalties.

For a feasible family A, define its indicator cost

$$I_A(E) = \begin{cases} 0 & \text{when } E \in A, \\ +\infty & \text{when } E \notin A. \end{cases}$$

Then the constrained problem becomes

$$E^{\bigstar} \in \arg\min \{D_{\pi}(E) + \lambda K_{\pi}(E) + I_{A\pi}(E) : E \in Eq(X)\}.$$

This unifies hard and soft formulations.

A finite penalty discourages violation.

An infinite penalty prohibits it.

The distinction is ontologically significant.

A hard transformation constraint says that a proposed object is invalid if its transformed image depends on the representative.

A soft transformation penalty says that small representative dependence may be tolerated.

A hard descent constraint says that no local object distinction may be erased globally.

A soft descent penalty says that some local distinctions may be revised.

A hard observable constraint says that selected outputs must be reproduced exactly.

A soft distortion term permits approximation.

Constraints can also be hierarchical.

One may require exact logical consistency first, exact safety preservation second, and then optimize prediction and complexity.

This can be represented lexicographically rather than through one weighted sum.

A lexicographic optimization might proceed as follows.

First, minimize the number of hard violations.

Second, among policies with minimum violation, minimize safety distortion.

Third, among those, minimize predictive distortion.

Fourth, minimize complexity.

Such a formulation prevents a large complexity gain from compensating for one unacceptable logical or safety failure.

Weighted sums assume that all costs can be exchanged numerically.

That assumption may be inappropriate when some constraints are noncompensable.

The choice between constrained, penalized, and lexicographic forms is therefore part of the admissibility policy π.

The mathematics should expose that choice rather than hide it.

7.3 Lagrange Multipliers for Object Formation

A constrained object-selection problem can often be expressed through a Lagrangian-like functional.

Suppose the goal is to minimize complexity subject to a distortion tolerance:

minimize $K\pi(E)$

subject to

$D\pi(E) \leq \varepsilon$.

Introduce a multiplier $\lambda \geq 0$ and define

$$L(E, \lambda) = K\pi(E) + \lambda(D\pi(E) - \varepsilon).$$

For a fixed λ, minimizing L over E favors policies balancing complexity against violation of the distortion bound.

Similarly, if closure and invariance are represented by nonnegative defect functions, define

$$L(E, \lambda, \mu, \nu) = K\pi(E) + \lambda(D\pi(E) - \varepsilon) + \mu L_{des}(E) + \nu S_{inv}(E).$$

The multipliers act as prices.

λ is the price of task distortion.

μ is the price of descent incoherence.

ν is the price of transformation instability.

When a constraint is expensive, policies violating it become unattractive.

This interpretation is useful in adaptive systems.

If a learned ontology repeatedly produces prediction errors, the effective distortion price can increase, encouraging refinement.

If computational resources become scarce, the complexity price can increase, encouraging coarsening.

If distributed observers begin to disagree, the descent price can increase.

If quotient dynamics become unstable, the invariance price can increase.

The ontology then changes in response to the costs imposed by its environment and task.

Classical multiplier theory is most direct in smooth convex optimization.

The space Eq(X) is discrete and generally nonconvex.

Consequently, standard conclusions about strong duality, differentiability, or unique multiplier values do not automatically apply.

The multiplier formulation remains useful as scalarization and constraint pricing, but it must not be interpreted as a guarantee that every constrained optimum is recovered by some weighted sum.

In finite discrete systems, one can evaluate each candidate policy and compare objective values directly.

The multiplier λ selects policies according to lines of the form

$J\lambda(E) = D\pi(E) + \lambda K\pi(E)$.

For two policies E and F, they have equal cost when

$D\pi(E) + \lambda K\pi(E)$

$D\pi(F) + \lambda K\pi(F)$.

If $K\pi(E) \neq K\pi(F)$, the exchange value is

$$\lambda EF = \frac{[D\pi(F) - D\pi(E)]}{[K\pi(E) - K\pi(F)]}.$$

At this critical λ, the two policies tie.

On one side of the critical value, the lower-distortion policy may be preferred.

On the other side, the lower-complexity policy may be preferred.

These exchange values generate the phase transitions studied later.

Multipliers therefore have both computational and conceptual meaning.

They express the current exchange rate between fidelity and simplicity.

There is no universal λ independent of purpose.

A high-stakes medical decision may assign an enormous price to distortion.

A low-bandwidth communication system may assign a large price to complexity.

An exploratory scientific model may vary λ deliberately to reveal stable resolution regimes.

The multiplier should therefore be reported as part of the object-selection result.

An ontology without its cost regime is incomplete.

7.4 The Objecthood Functional

The elementary objective

$$D\pi(E) + \lambda K\pi(E)$$

captures the central compression-distortion tradeoff.

A complete objecthood functional may include additional terms.

Define

$$\Phi\pi(E) = D\pi(E) + \lambda K\pi(E) + \mu L_{des}(E) + \nu S_{inv}(E) + \rho L_{sep}(E) + \sigma L_{model}(E) + \tau L_{repair}(E).$$

Here:

$D\pi(E)$ measures task distortion,

$K_\pi(E)$ measures quotient complexity,

$L_{des}(E)$ measures local-to-global incoherence,

$S_{inv}(E)$ measures transformation instability,

$L_{sep}(E)$ measures protected-distinction violations,

$L_{model}(E)$ measures mismatch between quotient behavior and the selected model family,

and $L_{repair}(E)$ measures deviation from a prior or supplied policy.

The selected policy is

$$E^\bigstar \in \arg\min \{\Phi_\pi(E) : E \in Eq(X)\}.$$

This functional should not be expanded with terms merely for completeness.

Every term must correspond to a real scientific, mathematical, or operational requirement.

An overcomplicated functional can make the theory unfalsifiable because any desired policy can be selected by adjusting enough weights.

The preferred construction is to begin with the smallest adequate objective.

Hard structural requirements should be imposed directly when possible.

Only uncertain or genuinely tradeable requirements should become penalties.

A useful decomposition is

$\Phi_\pi(E) =$
fidelity cost
+
representation cost
+
coherence cost
+
revision cost.

Fidelity cost measures failure to preserve the task.

Representation cost measures the burden of the quotient.

Coherence cost measures failure of descent and transformation laws.

Revision cost measures how much the new ontology differs from an established one.

Revision cost is important in adaptive systems.

An ontology that changes constantly may be numerically optimal at each moment but operationally unusable.

Let Eold be the previous identity policy.

Define a policy-change distance

dEq(E, Eold).

A simple finite measure counts pairs whose identity status changes:

dEq(E, F) = |E symmetric difference F|.

A class-sensitive measure can count object merges and splits.

Then include

τdEq(E, Eold)

in the functional.

The parameter τ introduces ontological inertia.

A new policy must improve distortion and complexity enough to justify the cost of changing concepts.

This can produce hysteresis.

The policy selected while λ increases may differ from the policy selected while λ decreases because revision itself has cost.

The objecthood functional may also be expressed as a benefit rather than a cost.

Define:

Cπ(E) = compression benefit,

Pπ(E) = predictive benefit,

Sπ(E) = stability benefit,

Hπ(E) = coherence benefit.

Then maximize

Vπ(E) =
Cπ(E)

+

$P\pi(E)$

+

$S\pi(E)$

+

$H\pi(E)$

–

costs.

The minimization and maximization forms are mathematically interchangeable after sign conventions.

The cost form is generally clearer because zero distortion and zero defect have direct interpretations.

The functional should satisfy several desirable properties where possible.

It should be invariant under admissible relabelings.

It should be computable or estimable from available evidence.

Its zero-defect conditions should correspond to exact structural laws.

Its complexity terms should be monotone under appropriate coarsening.

Its distortion terms should be monotone under appropriate refinement.

It should provide interpretable certificates explaining why one policy outranks another.

These properties make the optimization scientifically auditable.

The output should not be only the selected partition.

It should include the decomposition of $\Phi\pi(E\bigstar)$ into its constituent costs.

7.5 Least-Complexity Admissible Quotients

An important special case occurs when the relevant structures must be preserved exactly.

Let Aexact be the family of policies satisfying:

$D\pi(E) = 0$,

$C\mathcal{U}(E) = E$,

$E \in Inv(G)$,

and all protected distinctions remain separate.

The least-complexity admissible quotient problem is

$E^{\bigstar} \in \arg\min \{K_\pi(E) : E \in A_{exact}\}$.

This formulation asks:

Among all perfectly lawful identity systems, which one is simplest?

If K_π is monotone under coarsening, the problem seeks a maximally coarse feasible relation.

A greatest feasible relation, if it exists, is automatically optimal for every strictly monotone quotient-complexity measure.

However, a greatest feasible relation need not exist.

The feasible family may contain several incomparable maximal policies.

Each may compress different distinctions while preserving all requirements.

Their joins may violate a constraint.

Thus least-complexity admissible objecthood can be nonunique even under exact preservation.

If quotient cardinality is the complexity measure, several distinct policies may have the same minimum number of classes.

They may produce different object interpretations.

Additional criteria may select among them.

One may prefer:

the policy with the shortest description,

the policy with the simplest quotient dynamics,

the policy closest to a prior ontology,

the policy most robust to perturbation,

or the finest policy among equal-cardinality minimizers.

The exact problem has a useful universal form when the requirements are represented by observable kernels.

Let H be a family of observables to preserve exactly.

The coarsest policy preserving H alone is

$E_H = \bigcap_{h \in H} \mathrm{Ker}(h)$.

If EH is transformation-invariant and closure-coherent, it is the greatest exact feasible relation under those requirements.

It therefore produces the least-complexity exact quotient for every refinement-monotone Kπ.

If EH fails invariance or descent, it must be refined or the constraints must be reformulated.

This result shows how exact object selection can sometimes be solved algebraically without searching the full lattice.

Find the coarsest equivalence compatible with all required maps and operations.

The quotient by that relation is the minimal sufficient object system.

In automata theory, analogous constructions identify behaviorally indistinguishable states.

In statistics, sufficient statistics retain exactly the information required for a parameter or prediction.

In algebra, quotienting by a congruence preserves operations.

Variational objecthood unifies these patterns as least-complexity lawful identity.

The approximate theory extends the same principle when no exact sufficient quotient is available or when exact preservation is unnecessarily expensive.

7.6 Maximum-Compression Formulations

The variational principle can be written in terms of compression gain rather than complexity cost.

Let Kbase be the complexity of the unquotiented or baseline representation.

Define compression gain

$C(E) = K_{base} - K_\pi(E)$.

A maximum-compression formulation is

maximize $C(E)$

subject to

$D_\pi(E) \le \varepsilon$,

$L_{des}(E) = 0$,

$S_{inv}(E) = 0$,

and all protected distinctions are preserved.

Equivalently,

maximize C(E)

over $E \in A_{\pi,\varepsilon}$.

This formulation emphasizes the constructive effect of quotienting.

The optimal policy discards the largest amount of unnecessary detail without violating the fidelity threshold.

A benefit-cost form is

maximize $C(E) - \beta D_\pi(E)$.

If $C(E) = K_{base} - K_\pi(E)$, this differs from minimizing

$D_\pi(E) + \lambda K_\pi(E)$

only by scaling and an additive constant.

The compression formulation is often intuitive in applications.

A model-reduction problem asks how much state complexity can be removed.

A communication problem asks how many bits can be saved.

A scientific abstraction asks how many microscopic distinctions can be ignored.

An AI representation asks how compactly task-relevant structure can be encoded.

However, “maximum compression” can be misleading if K_π measures only quotient cardinality.

A small quotient may require a highly complex identity rule or highly complex induced dynamics.

The compression benefit should therefore be based on total effective complexity.

For example,

$C_{total}(E) =$
$K_{original}$
$-$
$[K_{policy}(E) + K_{objects}(E) + K_{dynamics}(E)]$.

A quotient is beneficial only when the total compressed description is smaller than the original.

This prevents apparent compression that merely transfers complexity from the state set into the identity rule or transition model.

One may also define local compression gains.

For a distributed policy family $\mathcal{E}$,

Clocal($\mathcal{E}$) =
Σi wi[Ki(original) − Ki(Ei)].

The global compression gain may differ from the sum of local gains because local objects can be identified across domains.

A global quotient can reduce duplication.

It can also create additional modeling complexity or descent defects.

The complete objective should therefore distinguish local and global compression.

Maximum-compression formulations naturally produce a resolution curve.

For each distortion tolerance ε, define

Kmin(ε) =
min {Kπ(E) : Dπ(E) ≤ ε and E satisfies structural constraints}.

The function Kmin(ε) is nonincreasing.

Allowing more distortion cannot increase the minimum achievable complexity.

Its discontinuities mark changes in the optimal quotient ontology.

This is the objecthood analogue of a rate-distortion curve.

7.7 Equivalent Pareto Formulations

Distortion and complexity are distinct objectives.

Rather than combining them immediately into one scalar, one may study them as a multiobjective optimization problem.

Associate to each feasible policy E the pair

(Dπ(E), Kπ(E)).

A policy E is dominated by F when

Dπ(F) ≤ Dπ(E),

$K\pi(F) \leq K\pi(E)$,

and at least one inequality is strict.

A dominated policy is inferior in both fidelity and simplicity.

It has no reason to be selected under a policy valuing both lower distortion and lower complexity.

A policy is Pareto-optimal when no feasible policy dominates it.

The set of Pareto-optimal policies forms the objecthood frontier.

This frontier contains the meaningful tradeoffs.

At one end lie fine policies with low distortion and high complexity.

At the other lie coarse policies with high distortion and low complexity.

Intermediate points represent different balances.

The weighted objective

$D\pi(E) + \lambda K\pi(E)$

selects supported points on the frontier.

Because $Eq(X)$ is discrete and the attainable cost set may be nonconvex, not every Pareto-optimal policy must be optimal for some λ.

Some nondominated policies may lie in concave portions of the discrete frontier and never minimize a simple weighted sum.

This is why the constrained formulations

minimize $K\pi(E)$ subject to $D\pi(E) \leq \varepsilon$

or

minimize $D\pi(E)$ subject to $K\pi(E) \leq B$

are also necessary.

By varying ε or B, one can recover Pareto policies missed by linear scalarization.

The Pareto viewpoint is especially useful when no principled exchange rate λ is known.

Instead of forcing one numerical tradeoff, the theory presents the full family of nondominated object systems.

A scientist or decision-maker can then examine:

which distinctions are lost at each policy,

which quotient dynamics arise,

which local observers are affected,

and how much complexity is saved.

The frontier may contain several policies with identical distortion and complexity values but different partitions.

These are cost-degenerate ontologies.

They are indistinguishable to the selected objective but may differ structurally or semantically.

This indicates that the objective is incomplete for selecting among them.

Additional criteria may be introduced, or the nonuniqueness may be retained as genuine underdetermination.

The Pareto formulation therefore prevents overinterpretation of one scalar optimum.

It reveals the range of equally rational conceptual resolutions under different tradeoffs.

A robust object policy may be characterized by occupying a broad or isolated region of the frontier.

If small changes in distortion tolerance or complexity budget leave the same policy optimal, the ontology is stable.

If tiny changes select radically different partitions, objecthood is fragile.

This robustness will be developed through phase-transition analysis.

7.8 Interpretation as an Algebraic Least-Action Principle

The variational objecthood principle resembles a least-action principle.

In mechanics, a path is selected by extremizing an action functional subject to boundary and dynamical conditions.

In variational objecthood, an equivalence relation is selected by minimizing an objecthood functional subject to observational and transformation constraints.

The objects being varied are different.

Mechanical variation ranges over trajectories or fields.

Objecthood variation ranges over identity policies.

The functional is also different.

It is not a physical action unless a particular physical theory identifies it as such.

The phrase algebraic least-action principle should therefore be understood structurally rather than as a claim that $D\pi(E) + \lambda K\pi(E)$ is a universal physical action.

The analogy consists of four elements.

First, there is a configuration space.

Here the configuration space is Eq(X), or a constrained lattice of policy families.

Second, there is a functional.

Here the functional measures distortion, complexity, incoherence, and instability.

Third, there are admissibility constraints.

Here they include closure coherence, transformation invariance, observable preservation, and protected distinctions.

Fourth, an extremizer is selected.

Here the extremizer is the preferred identity policy $E\bigstar$.

The principle may be stated:

Effective objects are the equivalence classes of an identity policy that extremizes preservation and compression under admissibility constraints.

This is an algebraic variational law because the variable is an equivalence structure.

It selects not how an existing object moves, but what counts as an object.

The principle also resembles free-energy minimization.

Distortion acts like an energetic penalty for combining incompatible states.

Complexity acts like a pressure toward fewer effective degrees of freedom.

The multiplier λ controls the balance.

At different λ, different quotient ontologies become optimal.

This analogy should also be used cautiously. The theory does not require thermodynamic interpretation. Its value is in suggesting why abrupt object fusion and splitting can occur as parameters change.

The principle provides a possible mathematical account of emergence.

A macro-object appears when a coarser equivalence relation becomes variationally preferable.

The detailed states do not disappear.

Their distinctions become too costly to preserve relative to their task relevance.

An object splits when internal distortion becomes too large for the compression benefit to justify.

Emergence and differentiation are therefore transitions in the optimizing identity policy.

The least-action interpretation also clarifies the status of objectivity.

The selected objects are not arbitrary names.

They minimize a stated functional under explicit constraints.

Different task policies can produce different optima, but the dependence is mathematically traceable.

One can identify which change in observable, transformation family, constraint, or multiplier caused the ontology to change.

This is stronger than saying merely that objects are observer-dependent.

They are policy-indexed extremizers.

The policy determines what must be preserved.

The variational law determines the best quotient under that policy.

Deeply robust objects may appear as minimizers across many policies, transformations, and parameter ranges.

Such objects are not merely convenient for one narrow task.

They occupy stable regions of the objecthood landscape.

The complete variational principle can now be stated in compact form.

For exact objecthood:

$E^\star \in \arg\min \{ D_\pi(E) + \lambda K_\pi(E) : E \in \mathrm{Fix}(C\mathcal{U}) \cap \mathrm{Inv}(G) \cap \mathrm{Pres}(H) \cap \mathrm{Sep}(P) \}.$

For approximate objecthood:

$$E^\star \in \arg\min \{ D_\pi(E) + \lambda K_\pi(E) + \mu L_{des}(E) + \nu S_{inv}(E) + \rho L_{sep}(E) : E \in Eq(X) \}.$$

For exact least-complexity objecthood:

$$E^\star \in \arg\min \{ K_\pi(E) : D_\pi(E) = 0 \text{ and } E \text{ satisfies all structural constraints} \}.$$

For Pareto objecthood:

select E from the nondominated frontier of

$(D_\pi(E), K_\pi(E))$

over the structural feasible region.

These formulations provide one unified theory of object selection.

The central conclusions of this section are:

Object formation is an optimization over identity policies.

The candidate space is Eq(X).

The structural feasible region is defined by closure, transformation, preservation, and separation constraints.

Distortion and complexity determine the primary tradeoff.

Hard constraints and finite penalties express different levels of admissibility.

Multipliers act as prices for distortion, complexity, incoherence, and instability.

Least-complexity quotients are maximally compressed policies preserving all required structure.

Maximum-compression and minimum-cost formulations express the same principle under different parameterizations.

Pareto analysis reveals tradeoffs not captured by one weighted sum.

The variational principle is an algebraic analogue of least action because it selects the identity structure itself.

The next section establishes existence results for finite state spaces. It examines minimizers, nonuniqueness, finest and coarsest optimal policies, tie-breaking under refinement, and stability of the selected ontology as the tradeoff parameter changes.

8. Existence, Extremizers, and Finite Objecthood Theorems

A variational principle becomes mathematically useful only when its solutions can be shown to exist and their possible multiplicity can be understood.

The objecthood problem searches over equivalence relations rather than over ordinary numerical variables. Even when the state space X is finite, the search space Eq(X) is combinatorially large, discrete, and generally nonconvex. Standard arguments based on derivatives, smoothness, or convexity therefore do not apply directly.

Finite state spaces nevertheless provide a rigorous foundational case.

If X is finite, then Eq(X) is finite. Every feasible subset of Eq(X) is also finite. Any real-valued objecthood functional therefore attains a minimum whenever the feasible family is nonempty.

This gives immediate existence of optimal identity policies.

Existence does not imply uniqueness.

Several different equivalence relations may produce the same distortion, complexity, and total objective value. Their quotient ontologies may be incomparable under refinement. One policy

may merge states that another distinguishes, while the second merges a different collection of states.

The optimization may therefore select a set of optimal policies rather than one policy.

The structure of this minimizer set matters.

There may be a unique finest optimum.

There may be a unique coarsest optimum.

There may be several refinement-minimal or refinement-maximal optima.

There may be no least or greatest optimum under refinement.

Additional tie-breaking principles may then be required.

The tradeoff parameter λ introduces a second layer of structure. For each fixed policy E, the objective

$J\lambda(E) = D\pi(E) + \lambda K\pi(E)$

is an affine function of λ.

The optimal value is the lower envelope of finitely many affine functions. It is therefore piecewise linear, continuous, and concave as a function of λ. Changes in the minimizing policy occur only at finitely many critical values unless some policies have identical cost lines.

Between critical values, the optimal policy family remains constant.

At a critical value, two or more ontologies can exchange optimality.

This provides an exact mathematical basis for phase transitions in conceptual resolution.

The finite theory also reveals an important monotonicity.

As λ increases, optimal complexity cannot increase.

The selected ontology must become no more complex, although it need not move monotonically upward in the refinement lattice. It may reorganize between incomparable partitions while still reducing total complexity.

Correspondingly, optimal distortion cannot decrease as greater weight is placed on simplicity.

The compression-distortion path is monotone in objective coordinates even when the identity policies themselves do not form a simple refinement chain.

This section proves these claims and states a finite variational objecthood theorem.

8.1 Existence on Finite State Spaces

Let X be a finite state space with

$|X| = n$.

The collection Eq(X) contains one equivalence relation for every partition of X.

The number of elements of Eq(X) is the nth Bell number.

Although this number may be extremely large, it is finite.

Let $A\pi$ be a feasible family of identity policies:

$A\pi \subseteq Eq(X)$.

The feasible family may impose:

transformation invariance,

closure coherence,

protected distinctions,

observable preservation,

cardinality bounds,

or other structural requirements.

Let

$J: A\pi \to \mathbb{R}$

be a real-valued objecthood functional.

The most common form is

$J(E) = D\pi(E) + \lambda K\pi(E)$,

but the result applies to any real-valued function on $A\pi$.

Finite Existence Proposition.

If X is finite and $A\pi$ is nonempty, then there exists at least one $E\bigstar \in A\pi$ such that

$J(E\bigstar) \leq J(E)$

for every $E \in A\pi$.

Proof.

Because Eq(X) is finite, Aπ is finite.

The set of objective values

{J(E) : E ∈ Aπ}

is a finite nonempty subset of $\mathbb{R}$.

Every finite nonempty subset of $\mathbb{R}$ has a minimum.

Choose E★ attaining that minimum.

Therefore E★ is an optimal identity policy.

The proposition is elementary, but it establishes a complete existence theorem for finite object discovery.

No differentiability is required.

No convexity is required.

No continuity is required.

No lattice monotonicity is required.

The only essential assumptions are:

a finite state space,

a nonempty feasible family,

and a finite real objective value for every feasible policy.

The same conclusion holds for a penalized formulation over all of Eq(X):

E★ ∈ arg min {Φπ(E) : E ∈ Eq(X)}.

Because Eq(X) is finite, every real-valued penalized functional Φπ attains a minimum.

Hard constraints require separate feasibility.

If Aπ is empty, no exact solution exists.

A penalized formulation still produces a minimizer, but that minimizer may violate one or more structural conditions.

The distinction must be preserved.

A minimum of a penalized objective is not automatically a lawful object system.

It is the least costly compromise under the selected penalties.

An exact finite objecthood problem therefore has three possible outcomes.

First, the feasible family is nonempty and a unique minimizer exists.

Second, the feasible family is nonempty and several minimizers exist.

Third, the feasible family is empty.

The third outcome is not a failure of optimization.

It is a proof that the structural requirements are jointly inconsistent.

The system must then revise its constraints, observations, transformations, or protected distinctions.

Finite existence also applies to distributed policy families.

Let

L$\mathcal{U}$ = ∏i∈I Eq(Ui)

for a finite family of finite domains Ui.

The product lattice L$\mathcal{U}$ is finite.

Any nonempty feasible subset of L$\mathcal{U}$ therefore supports at least one minimum of every real-valued distributed objecthood functional.

Thus local-to-global object selection has exact minimizers in every finite observer network.

8.2 Compactness of Finite Policy Lattices

Finite existence can also be expressed through compactness.

Equip Eq(X) with the discrete topology.

Every subset is open.

Because Eq(X) is finite, it is compact.

Every sequence of equivalence relations contains a constant subsequence.

Every real-valued function on Eq(X) is continuous in the discrete topology.

Therefore every real-valued function attains both a minimum and a maximum.

This topological language is not necessary for the finite theorem, but it prepares the extension to infinite state spaces.

In continuous variational problems, existence is often established by:

compactness of the admissible space,

lower semicontinuity of the objective,

and closure of the constraint set.

The finite lattice satisfies these properties automatically.

The refinement order supplies an additional structure.

Eq(X) is a finite complete lattice.

Every subset has refinement-minimal and refinement-maximal elements, although it need not have a unique least or greatest element.

Every ascending or descending chain terminates.

No infinite strict refinement sequence exists.

No infinite strict coarsening sequence exists.

Consequently, any iterative algorithm that strictly improves the objective at every accepted policy change must terminate after finitely many steps.

Termination does not imply global optimality.

A merge-split algorithm may stop at a local minimum relative to its allowed moves.

The finite lattice prevents endless strict improvement, but it does not prevent convergence to a suboptimal policy.

The complete lattice structure also permits order-based compactness arguments.

Every sequence of monotone refinements stabilizes.

Every sequence of monotone coarsenings stabilizes.

Closure iteration stabilizes after finitely many strict changes.

Invariant-refinement procedures stabilize.

Alternating closure operators on a finite lattice eventually reach a fixed point or a cycle. If the combined update is monotone and extensive, the sequence stabilizes at a fixed point.

These finite stabilization results make exact computation conceptually straightforward even when it is combinatorially expensive.

The policy lattice therefore plays three roles.

It is the domain of optimization.

It supplies the refinement geometry of possible ontology changes.

It guarantees finite stabilization of monotone policy operations.

8.3 Existence of Minimizers

Define the minimizer set

$$\mathrm{Min}_{\pi,\lambda} = \{E \in A_\pi : J_{\pi,\lambda}(E) = \min_{F \in A_\pi} J_{\pi,\lambda}(F)\}.$$

When X is finite and A_π is nonempty,

$$\mathrm{Min}_{\pi,\lambda} \neq \emptyset.$$

The minimizer set is itself finite.

It may contain one or many policies.

Several common formulations therefore have exact finite solutions.

Penalized objecthood:

$$E^\star \in \arg\min \{ D_\pi(E) + \lambda K_\pi(E) + \mu L_{des}(E) + \nu S_{inv}(E) : E \in Eq(X) \}.$$

If every term is finite-valued, a minimizer exists.

Exact constrained objecthood:

$$E^\star \in \arg\min \{ D_\pi(E) + \lambda K_\pi(E) :$$

$E \in \mathrm{Fix}(C\mathcal{U}) \cap \mathrm{Inv}(G) \cap \mathrm{Pres}(H) \cap \mathrm{Sep}(P)$
}.

A minimizer exists whenever the intersection is nonempty.

Minimum-complexity objecthood:

$E\bigstar \in \arg\min \{$
$K\pi(E) :$
$E \in A\pi$
and
$D\pi(E) \leq \varepsilon$
}.

A minimizer exists whenever at least one feasible policy satisfies the distortion bound.

Minimum-distortion objecthood:

$E\bigstar \in \arg\min \{$
$D\pi(E) :$
$E \in A\pi$
and
$K\pi(E) \leq B$
}.

Again, existence follows from nonempty finite feasibility.

Exact zero-distortion objecthood:

$E\bigstar \in \arg\min \{$
$K\pi(E) :$
$E \in A\pi$
and
$D\pi(E) = 0$
}.

A solution exists when the zero-distortion feasible family is nonempty.

Pareto objecthood also has finite solutions.

Because the feasible set is finite, at least one nondominated policy exists.

To see this, begin with any feasible policy. If it is dominated, move to a dominating policy. Because the feasible family is finite and at least one objective improves strictly at each move, the process terminates at a nondominated policy.

Thus the compression-distortion frontier is nonempty whenever Aπ is nonempty.

The finite theory therefore guarantees existence across all principal formulations.

The substantive problem is not whether some optimum exists.

It is whether the optimum is unique, scientifically meaningful, robust, and computationally accessible.

8.4 Nonuniqueness of Optimal Policies

An optimal identity policy need not be unique.

Let E and F be distinct feasible equivalence relations.

They tie at parameter λ when

Dπ(E) + λKπ(E)

Dπ(F) + λKπ(F).

Several forms of nonuniqueness are possible.

Cost-equivalent policies may have identical distortion and complexity:

Dπ(E) = Dπ(F),

Kπ(E) = Kπ(F).

Then they tie for every λ.

The selected objective cannot distinguish them.

A second form occurs at one critical λ.

The policy E has lower distortion but greater complexity.

The policy F has greater distortion but lower complexity.

At a specific exchange value, the two costs are equal.

A third form occurs when several policies produce the same minimum despite differing in all component costs because additional penalties compensate.

A fourth form occurs through symmetry.

If a task and state space are invariant under a transformation that carries E to F, the two policies may be structurally equivalent and necessarily have equal cost.

A fifth form occurs through incomplete objectives.

Two policies may preserve different scientifically meaningful structures while the selected Dπ and Kπ assign them the same scalar values.

Nonuniqueness can represent several different situations.

It may represent genuine symmetry.

It may represent observational underdetermination.

It may indicate that the task does not care which ontology is selected.

It may indicate that the objective is too coarse.

It may represent a phase-transition point where two conceptual resolutions are equally favored.

These interpretations should not be conflated.

The output of the variational theory should therefore include the complete minimizer set when computationally feasible, or at least a description of alternative minimizers.

Reporting one arbitrarily chosen optimizer can create a false impression of uniqueness.

Suppose E and F are incomparable under refinement.

Then E preserves some distinctions that F erases, while F preserves other distinctions that E erases.

Equal objective value does not imply semantic equivalence.

The quotients X/E and X/F may support different concepts.

One may be preferable for explanation.

The other may be preferable for intervention.

If the primary objective does not distinguish them, a secondary policy must decide whether one should be preferred.

Nonuniqueness is especially important in artificial intelligence.

A learning system may discover multiple equally efficient representations.

Forcing one canonical ontology may be unnecessary or harmful.

The system could retain several object systems, switch between them by task, or represent the minimizer family as an unresolved conceptual alternative.

The minimizer set is itself a mathematical object worthy of analysis.

One may study:

its refinement order,

its symmetries,

its shared identities,

its disputed identities,

and its stability under parameter changes.

Define the core identity relation of the minimizer family by

$E_{core} = \bigcap_{E \in Min_{\pi,\lambda}} E.$

The relation E_{core} contains exactly the identities accepted by every optimal policy.

It is an equivalence relation because intersections of equivalence relations are equivalence relations.

It is finer than every minimizer.

E_{core} need not itself be optimal.

Its complexity may be greater and its objective value higher.

Nevertheless, it represents the identities that are robust across all optima.

One may also form the generated envelope

$E_{envelope} = \bigvee_{E \in Min_{\pi,\lambda}} E.$

This is the least equivalence relation containing every identity accepted by any minimizer.

$E_{envelope}$ also need not be optimal or feasible under all constraints.

It represents the broadest identity structure suggested somewhere in the optimal family.

The interval between core and envelope describes ontological uncertainty.

8.5 Finest and Coarsest Minimizers

Because $Eq(X)$ is partially ordered by refinement, the minimizer set can be analyzed order-theoretically.

Let

$M = \mathrm{Min}_{\pi,\lambda}$.

A minimizer $E \in M$ is refinement-minimal in M when there is no distinct $F \in M$ satisfying

$F \preceq E$.

Because $F \preceq E$ means F is finer, a refinement-minimal minimizer cannot be replaced by a strictly finer optimum.

It is locally among the finest optimal policies.

A minimizer $E \in M$ is refinement-maximal in M when there is no distinct $F \in M$ satisfying

$E \preceq F$.

It cannot be replaced by a strictly coarser optimum.

It is locally among the coarsest optimal policies.

Every nonempty finite partially ordered set has at least one minimal element and at least one maximal element.

Therefore every finite minimizer set contains at least one refinement-minimal and one refinement-maximal policy.

There need not be only one of either.

A finest minimizer in the strong sense is a least element $E_{\min} \in M$ satisfying

$E_{\min} \preceq E$

for every $E \in M$.

If it exists, it is unique.

It preserves at least every distinction preserved by any other optimal policy.

A coarsest minimizer in the strong sense is a greatest element $E_{\max} \in M$ satisfying

$E \preceq E_{\max}$

for every $E \in M$.

If it exists, it is unique.

It merges at least every pair merged by any other optimal policy.

The finest and coarsest minimizers need not exist.

The minimizer set may contain incomparable policies with no optimal common refinement or common coarsening.

The meet of all minimizers exists in Eq(X):

$\bigwedge M = \bigcap_{E \in M} E$.

But this meet may not belong to M.

It can have lower distortion and greater complexity than the tied optimum.

Similarly, the join

$\bigvee M = \mathrm{EqCl}(\bigcup_{E \in M} E)$

may fail to be optimal or feasible.

Thus the lattice operations of Eq(X) do not automatically remain inside the minimizer set.

Under additional assumptions, closure may occur.

If the objective and feasible family are meet-compatible, the meet of minimizers may remain a minimizer.

If they are join-compatible, the join may remain a minimizer.

Such conditions are special and should be proved for each application rather than assumed.

The distinction between minimal and least, and between maximal and greatest, is essential.

A finite set always has minimal and maximal elements.

It need not have a unique finest or coarsest element.

This matters when describing the solution.

One should not claim “the finest optimum” merely because one refinement-minimal optimum has been found.

8.6 Tie-Breaking by Refinement

When the primary objecthood functional has several minimizers, a secondary principle can select among them.

One possible rule is finest-optimum selection.

Among all primary minimizers, choose a policy preserving the greatest amount of detail.

A cardinal version is

maximize |X/E|

subject to

E ∈ Minπ,λ.

A relation-size version is

minimize |E|

subject to

E ∈ Minπ,λ,

where |E| counts ordered equivalent pairs.

Finer policies generally contain fewer related pairs.

This rule is conservative.

When several ontologies have equal primary cost, it avoids unnecessary identity claims.

It preserves distinctions unless their merger produces a strict improvement.

A second rule is coarsest-optimum selection.

Among all primary minimizers, choose a policy with the greatest compression:

minimize |X/E|

subject to

E ∈ Minπ,λ.

Equivalently, maximize relation size when appropriate.

This rule prefers broader abstraction when the primary functional is indifferent.

A third rule minimizes distance to a prior policy E0:

minimize dEq(E, E0)

subject to

E ∈ Minπ,λ.

This stabilizes conceptual continuity.

A fourth rule selects the policy with the shortest description.

A fifth selects the simplest quotient dynamics.

A sixth selects the policy most robust under perturbations.

A seventh retains all minimizers rather than selecting one.

Tie-breaking can be expressed lexicographically.

For example:

first minimize $J_{\pi,\lambda}(E)$,

then among the minimizers minimize $K_{desc}(E)$,

then among remaining ties minimize $d_{Eq}(E, E0)$,

then among remaining ties select the finest policy.

Lexicographic optimization prevents a secondary criterion from overturning even a small advantage in the primary objective.

An alternative is to add a small regularization term:

$J_{\varepsilon}(E) =$
$J_{\pi,\lambda}(E)$
$+$
$\varepsilon R(E)$,

where $\varepsilon > 0$ is small and R is the tie-breaking cost.

If ε is sufficiently small relative to the objective gap between primary minimizers and nonminimizers, the regularizer selects among the primary minimizers without changing which primary value is optimal.

In a finite problem, let

$\gamma =$
minimum positive difference between a nonoptimal objective value and the optimum.

If R is bounded and ε is chosen so that its maximum possible contribution is smaller than γ, primary optimality is preserved.

Tie-breaking by refinement has a philosophical interpretation.

Finest-optimum selection follows a principle of identity restraint:

Do not identify states unless doing so is necessary for optimality.

Coarsest-optimum selection follows a principle of maximal abstraction:

Do not preserve distinctions unless doing so is necessary for optimality.

Neither rule is universally correct.

The first prioritizes epistemic caution.

The second prioritizes compression.

The choice should be included in policy π.

A system that silently applies one rule may embed a substantial ontological preference without acknowledging it.

8.7 Stability Under Perturbation of λ

Let the feasible family A_π be finite and independent of λ.

For every $E \in A_\pi$, define

$J_E(\lambda) = D_\pi(E) + \lambda K_\pi(E)$.

Each J_E is an affine function of λ.

Define the optimal value function

$V(\lambda) = \min_{E \in A_\pi} J_E(\lambda)$.

Because V is the pointwise minimum of finitely many affine functions, V is:

continuous,

piecewise linear,

and concave.

The concavity follows because a pointwise minimum of affine functions is concave.

The slope of an active affine segment is the complexity $K_\pi(E)$ of the policy optimal on that segment.

Critical values occur where two or more policy lines intersect.

For policies E and F with

$K_\pi(E) \neq K_\pi(F)$,

their exchange value is

$\lambda_{EF} =$
$[D_\pi(F) - D_\pi(E)]$

/
[Kπ(E) − Kπ(F)].

Only nonnegative exchange values are relevant when λ ≥ 0.

The set of critical values is finite because the feasible family is finite.

Some pairwise intersections may never appear on the lower envelope and therefore do not correspond to actual ontology changes.

Between consecutive active critical values, the minimizer set remains constant except for policies whose affine cost functions are identical.

If a unique policy E★ is optimal at λ0 and λ0 is not critical, then E★ remains optimal in an open interval around λ0.

This gives local parameter stability.

A quantitative bound can be obtained from the objective gap.

Let

γ(λ0) =
minF≠E★ [JF(λ0) − JE★(λ0)].

Because E★ is uniquely optimal,

γ(λ0) > 0.

Let

M =
maxF≠E★ |Kπ(F) − Kπ(E★)|.

If M > 0, then E★ remains uniquely optimal for every λ satisfying

|λ − λ0| < γ(λ0)/M.

A stricter bound may be used to avoid equality at the boundary.

If M = 0, changes in λ do not alter cost differences between E★ and the alternatives.

The objective gap therefore measures robustness.

A large gap supports a stable ontology.

A small gap indicates that minor changes in the complexity price may change the selected object system.

The optimal distortion and complexity obey monotonicity laws.

Let

$0 \leq \lambda 1 < \lambda 2$.

Choose any E1 optimal at λ1 and any E2 optimal at λ2.

Then

$K\pi(E2) \leq K\pi(E1)$,

and

$D\pi(E2) \geq D\pi(E1)$.

Proof.

Optimality gives

$$D\pi(E1) + \lambda 1 K\pi(E1) \leq D\pi(E2) + \lambda 1 K\pi(E2),$$

and

$$D\pi(E2) + \lambda 2 K\pi(E2) \leq D\pi(E1) + \lambda 2 K\pi(E1).$$

Adding and rearranging yields

$(\lambda 2 - \lambda 1)[K\pi(E2) - K\pi(E1)] \leq 0$.

Because $\lambda 2 - \lambda 1 > 0$,

$K\pi(E2) \leq K\pi(E1)$.

Substituting into the first optimality inequality gives

$D\pi(E1) \leq D\pi(E2)$.

Thus increasing the price of complexity cannot produce an optimum with greater complexity or lower distortion.

The selected equivalence relations need not satisfy

$E1 \preceq E2$.

The higher-λ policy may be incomparable with the lower-λ policy.

It may reorganize the ontology rather than merely merge existing classes.

What is monotone is the objective coordinate pair:

complexity weakly decreases,

distortion weakly increases.

Additional assumptions are required to obtain a nested chain of optimal partitions.

This distinction is important for interpreting conceptual phase transitions.

A change in λ may cause simple object fusion.

It may also replace one classification principle with another.

The number of objects may fall even while some previously merged states become separated and other states become fused.

Ontology can reorganize while becoming globally simpler.

Perturbations in the distortion or complexity estimates can be analyzed similarly.

If E★ has an objective gap γ and all estimated objective errors are smaller than γ/2, the identity of the unique optimum is preserved.

Thus finite objecthood has a natural statistical robustness criterion.

The selected ontology is trustworthy only when estimation uncertainty is small relative to the gap separating it from alternatives.

8.8 A Finite Variational Objecthood Theorem

The preceding results can be combined into one theorem.

Finite Variational Objecthood Theorem.

Let X be a finite nonempty state space.

Let Eq(X) be its lattice of equivalence relations.

Let Aπ ⊆ Eq(X) be a nonempty feasible family determined by any collection of exact constraints, including transformation invariance, closure coherence, observable preservation, and protected distinctions.

Let

$D\pi: A\pi \to \mathbb{R}\geq 0$

be a distortion functional.

Let

$K\pi: A\pi \to \mathbb{R}\geq 0$

be a complexity functional.

For $\lambda \geq 0$, define

$J\pi,\lambda(E) = D\pi(E) + \lambda K\pi(E)$.

Then the following statements hold.

First, at least one optimal policy exists:

$\arg\min \{J\pi,\lambda(E) : E \in A\pi\} \neq \emptyset$.

Second, the minimizer set is finite and contains at least one refinement-minimal and at least one refinement-maximal element.

Third, a unique finest or coarsest minimizer need not exist.

Fourth, if the minimizer is unique at $\lambda 0$, then it remains unique for all sufficiently small perturbations of λ.

Fifth, the optimal value function

$V(\lambda) = \min E \in A\pi\ J\pi,\lambda(E)$

is continuous, concave, and piecewise linear.

Sixth, changes in the minimizer family occur only at finitely many critical values of λ, apart from policies with identical affine cost functions.

Seventh, if

$0 \leq \lambda 1 < \lambda 2$,

and E1 and E2 are optimal at $\lambda 1$ and $\lambda 2$ respectively, then

$K\pi(E2) \leq K\pi(E1)$,

and

$D\pi(E2) \geq D\pi(E1)$.

Thus increased pressure for simplicity cannot increase optimal complexity or decrease optimal distortion.

Eighth, the optimal policies themselves need not form a refinement chain as λ varies.

Ninth, if several policies minimize the primary objective, a secondary refinement, description-length, dynamical, prior-distance, or robustness criterion can be applied lexicographically.

Tenth, if Aπ is empty, no exact identity policy satisfies the complete constraint system. Any penalized solution must then be interpreted as an approximate compromise rather than an exact objecthood solution.

The theorem supplies a rigorous finite foundation for object selection.

It says that optimized object systems exist whenever the exact requirements are jointly satisfiable.

It also prevents several overstatements.

The variational principle does not guarantee one ontology.

It does not guarantee that optimal policies are nested.

It does not guarantee that a weighted sum recovers every Pareto-optimal policy.

It does not guarantee that the computational search is easy.

What it guarantees is a well-defined finite extremization problem with exact solutions and a piecewise-stable parameter structure.

The theorem supports a practical object-discovery procedure.

First, specify X.

Second, define the structural feasible family Aπ.

Third, define Dπ and Kπ.

Fourth, enumerate, search, or approximate the feasible policy lattice.

Fifth, identify the minimizer set.

Sixth, calculate the objective gap.

Seventh, test stability across λ and estimation uncertainty.

Eighth, apply a stated tie-breaking rule only when necessary.

Ninth, report the selected quotient together with alternative minimizers and active constraints.

This procedure converts an informal choice of concepts into an auditable mathematical result.

The selected objects are the classes of an extremizing identity policy.

Their existence follows from finiteness and feasibility.

Their uniqueness depends on the discriminating power of the objective and constraints.

Their robustness depends on the gap between the optimum and competing policies.

Their resolution depends on the cost assigned to complexity.

The finite theorem therefore provides the first complete existence theory for variational objecthood.

The next section studies the critical values at which the optimizing policy changes. It interprets these changes as phase transitions in conceptual resolution, including abrupt object fusion, object birth, stable ontological plateaus, bifurcation, and hysteresis.

9. Phase Transitions in Conceptual Resolution

The finite variational theory shows that an optimal identity policy exists whenever the feasible family is nonempty. It also shows that the selected policy can change only at critical values of the complexity parameter λ.

These changes are not merely numerical adjustments.

When the optimizer changes, the quotient object system changes.

Classes may merge.

Classes may split.

One organizing principle may be replaced by another.

An object previously treated as fundamental may disappear into a larger class.

A formerly unified class may divide into several objects.

The effective ontology therefore undergoes transitions as the relative value of compression and fidelity changes.

This section develops those transitions as a phase theory of conceptual resolution.

The analogy with physical phase transitions is structural.

A physical phase is selected by minimizing a thermodynamic potential under external conditions.

An object ontology is selected by minimizing an objecthood functional under task, observation, and resource conditions.

A control parameter changes.

For a range of values, one structure remains optimal.

At a critical value, another structure acquires equal or lower cost.

The selected configuration changes, sometimes discontinuously.

The objecthood functional is not assumed to be a physical free energy. The analogy lies in extremization, parameter dependence, stability regimes, degeneracy, and abrupt structural reorganization.

The control parameter λ measures the price of complexity relative to distortion.

Other control parameters may also matter:

noise level,

time horizon,

intervention cost,

observer reliability,

descent penalty,

transformation family,

or available computational resources.

Changing any of these can alter the optimal identity policy.

The resulting phase structure reveals more than one best quotient. It reveals which object systems remain stable across ranges of conditions, which mergers occur at identifiable thresholds, and which concepts exist only under finely tuned assumptions.

A robust object appears as part of a policy that remains optimal over a substantial parameter region.

A fragile object appears only at a boundary or under a narrow weighting.

An emergent object appears when a coarser class becomes variationally preferable.

An object birth appears when a previously unified class becomes too internally distorted and splits.

Thus object formation and object dissolution can be treated as transitions in the optimizer of the identity-policy lattice.

9.1 The Compression-Distortion Frontier

Let $A\pi$ be the feasible family of identity policies.

Associate to each $E \in A\pi$ the point

$(D\pi(E), K\pi(E))$

in the distortion-complexity plane.

The horizontal coordinate measures task loss.

The vertical coordinate measures object-system complexity.

Policies toward the lower left are preferred because they have both lower distortion and lower complexity.

A policy E is dominated when there exists $F \in A\pi$ such that

$D\pi(F) \leq D\pi(E)$,

$K\pi(F) \leq K\pi(E)$,

and at least one inequality is strict.

A dominated policy is inferior in both primary objectives.

It cannot minimize

$D\pi(E) + \lambda K\pi(E)$

for any $\lambda > 0$.

The nondominated policies form the compression-distortion frontier.

This frontier is the set of meaningful objecthood tradeoffs.

At one end lies a fine policy with low distortion and high complexity.

At the other lies a coarse policy with low complexity and high distortion.

Intermediate policies exchange some fidelity for some compression.

The frontier need not be smooth.

For finite state spaces it is a finite set of attainable points.

Several policies may occupy the same point.

The frontier may contain gaps, corners, vertical alignments, horizontal alignments, and nonconvex segments.

A weighted objective selects a point minimizing

$D + \lambda K$.

Geometrically, lines of equal objective value satisfy

$D + \lambda K = \text{constant}$.

Their slope in the distortion-complexity plane is

$-1/\lambda$

when distortion is placed on the horizontal axis and complexity on the vertical axis.

Moving the line toward the lower-left feasible region identifies an optimal frontier point.

As λ increases, complexity becomes more expensive. The supporting line rotates so that lower-complexity policies become more favorable.

The selected point moves along the frontier toward lower K and usually higher D.

Not every Pareto point must be selected by a weighted sum.

A nondominated point lying in a nonconvex portion of the attainable region may never minimize $D + \lambda K$ for any λ.

Constrained formulations are therefore also required.

Define the minimum achievable complexity under distortion tolerance ε:

$$K_{\min}(\varepsilon) = \min \{K_\pi(E) : E \in A_\pi \text{ and } D_\pi(E) \leq \varepsilon\}.$$

Define the minimum achievable distortion under complexity budget B:

$$D_{\min}(B) = \min \{D_\pi(E) : E \in A_\pi \text{ and } K_\pi(E) \leq B\}.$$

These step functions recover all operationally relevant frontier points.

$K_{\min}(\varepsilon)$ is nonincreasing in ε.

Dmin(B) is nonincreasing in B.

Allowing more distortion cannot make the best achievable complexity worse.

Allowing more complexity cannot make the best achievable distortion worse.

The frontier has an ontological interpretation.

Each point represents a possible resolution of the world.

Moving toward lower distortion creates more distinctions.

Moving toward lower complexity erases more distinctions.

The frontier therefore traces the best available object systems across levels of permitted forgetting.

Policies away from the frontier are inefficient ontologies. They preserve unnecessary complexity without gaining fidelity, or they introduce unnecessary distortion without gaining compression.

The frontier is also task-relative.

Changing π changes the distortion values, feasible family, and possibly the complexity measure.

A policy lying on the frontier for prediction may be dominated for intervention.

A quotient optimal for short-horizon dynamics may be poor for long-horizon explanation.

Thus there is no universal compression-distortion frontier independent of purpose.

However, policies appearing on many frontiers may represent robust object structures.

9.2 Critical Values of the Tradeoff Parameter

For each feasible policy E, define its cost line

$J_E(\lambda) = D_\pi(E) + \lambda K_\pi(E)$.

The optimal value is

$V(\lambda) = \min \{J_E(\lambda) : E \in A_\pi\}$.

The function V is the lower envelope of finitely many affine lines.

A critical value is a value of λ at which the active minimizer set changes.

Suppose E and F have different complexities.

They tie when

$D_\pi(E) + \lambda K_\pi(E)$

$D\pi(F) + \lambda K\pi(F)$.

Solving gives

$$\lambda_{EF} = \frac{[D\pi(F) - D\pi(E)]}{[K\pi(E) - K\pi(F)]}.$$

This value is relevant only when $\lambda_{EF} \geq 0$ and the common cost lies on the lower envelope.

Many pairwise intersections occur above the optimum and do not produce actual transitions.

At a genuine critical value λ_c, at least two policies have equal minimum cost.

For λ just below λ_c, one policy may be preferred.

For λ just above λ_c, another may be preferred.

If E has lower distortion and greater complexity than F, then increasing λ can cause a transition from E to F.

The system sacrifices fidelity to obtain a simpler object space.

The jump in complexity is

$$\Delta K = K\pi(F) - K\pi(E),$$

which is nonpositive for an upward change in λ.

The jump in distortion is

$$\Delta D = D\pi(F) - D\pi(E),$$

which is nonnegative.

The transition is discontinuous at the policy level even though the optimal value $V(\lambda)$ remains continuous.

This distinction parallels first-order phase transitions.

The minimized potential changes continuously, while the selected state or an order parameter jumps.

A natural order parameter for conceptual resolution is

$$N_{obj}(E) = |X/E|$$

for finite X.

Other order parameters include:

quotient entropy,

average class size,

maximum class size,

number of protected distinctions retained,

or dimension of an effective representation.

At a critical λ, Nobj may drop abruptly.

Several objects fuse at once.

Alternatively, the class count may remain unchanged while the partition reorganizes.

The transition then changes object identity without changing object number.

This is an ontological transition invisible to cardinal complexity alone.

The critical value has a direct exchange interpretation.

Suppose F is coarser than E.

The move from E to F saves

$K\pi(E) - K\pi(F)$

units of complexity and incurs

$D\pi(F) - D\pi(E)$

units of distortion.

The transition occurs when the price λ of complexity reaches the ratio

distortion increase divided by complexity saved.

Below that price, the added fidelity of E is worth its complexity.

Above it, the compression of F is worth its distortion.

Critical values therefore quantify the exact cost at which a distinction ceases to be worth preserving.

This is one of the most concrete outputs of the variational theory.

A concept merger does not occur merely because two classes are similar.

It occurs when the reduction in effective complexity outweighs the task loss under the selected exchange rate.

9.3 Sudden Object Fusion

Object fusion occurs when two or more classes of the current optimal policy become one class under a newly optimal coarser policy.

Let E be optimal below a critical value and F be optimal above it, with

$E \preceq F$.

Every F-class is a union of E-classes.

Suppose

C1, C2, …, Cm

are distinct E-classes contained in one F-class C.

At the transition, the m objects become one.

The fusion may be local.

Only one pair of classes merges, while the rest of the ontology remains unchanged.

It may also be collective.

A single policy change can merge many classes at once.

Collective fusion occurs when the objective favors a coarser equivalence relation whose transitive closure connects a large portion of the state space.

The distortion increase produced by fusion depends on the internal disagreement among the merged classes.

Let

$\Delta D_{fusion} =$
$D_\pi(F) - D_\pi(E)$.

Let

$\Delta K_{fusion} =$
$K_\pi(E) - K_\pi(F)$.

Fusion becomes favorable when

$\Delta D_{fusion} < \lambda \Delta K_{fusion}$.

At equality, the two policies tie.

This condition is the local economic law of object merger.

The complexity saved by treating several objects as one exceeds the task cost of losing their distinction.

Fusion can represent several phenomena.

In perception, several views become one recognized object.

In physics, many microstates become one macrostate.

In biology, several measured variants become one functional type.

In mathematics, several presentations become one isomorphism class.

In AI, several clusters become one concept.

In distributed observation, several local object traces become one global object.

Not every fusion is emergence.

Some fusions are destructive.

A quotient may merge states with incompatible intervention responses or unstable dynamics.

Such a merger has high distortion or violates a hard constraint.

Emergent fusion requires that the new class remain lawful relative to the selected task.

The transition can also reveal scale dependence.

At low λ, the system preserves fine-grained objects.

At high λ, it operates under resource pressure and forms broader categories.

The objects are not equally fundamental under every regime.

Their effective reality depends on the resolution at which lawful compression is optimized.

Fusion may occur through a sequence of critical values.

Classes merge hierarchically:

$E0 \preceq E1 \preceq E2 \preceq \cdots$.

This produces a dendrogram-like ontology.

However, such nesting is not guaranteed.

A higher-λ optimum can be incomparable with the lower-λ optimum.

The system may separate some previously fused states while merging others, provided total complexity decreases.

This is reclassification rather than pure aggregation.

Therefore object fusion should be diagnosed by comparing actual equivalence classes, not only class counts.

9.4 Sudden Object Birth

Object birth is the reverse structural event.

A previously unified class divides into two or more classes under a newly preferred finer policy.

When λ alone increases, the finite theorem shows that optimal complexity cannot increase. Therefore a pure increase in λ cannot force a more complex optimum.

Object birth can nevertheless occur under several conditions.

First, λ may decrease.

As complexity becomes less expensive, a finer policy may become optimal.

Second, the distortion functional may change.

New evidence can reveal previously hidden task differences inside a class.

Third, the task policy π may change.

A classification adequate for prediction may split under intervention requirements.

Fourth, the transformation family G may expand.

New admissible operations may expose instability inside a class.

Fifth, protected distinctions may be added.

A formerly legal merger may become infeasible.

Sixth, a descent defect may be discovered.

A global class may contain two locally distinct objects that must be separated.

Let F be the earlier coarse policy and E the later fine policy, with

$E \preceq F$.

Suppose one F-class C is partitioned into E-classes

C1, C2, …, Cm.

The single object C has given birth to m distinguishable objects.

The refinement increases complexity:

$K\pi(E) \geq K\pi(F)$.

It reduces distortion or restores feasibility:

$D\pi(E) \leq D\pi(F)$

under ordinary monotonicity assumptions.

The split becomes favorable when the reduction in distortion exceeds the increased complexity cost.

If the control parameter is λ decreasing, the transition occurs when preserving the distinction becomes affordable.

If the control parameter is a change in task loss, birth occurs when internal heterogeneity becomes sufficiently consequential.

Object birth therefore corresponds to the discovery that one class no longer supports lawful substitution.

The class had been treated as one object because its internal differences were irrelevant or unobserved.

New conditions make those differences matter.

Examples include:

a disease category dividing into treatment-responsive subtypes,

a species classification splitting after genomic analysis,

a physical phase dividing after an order parameter becomes detectable,

an AI concept splitting after repeated prediction errors,

and a global identity cluster separating after one bridge is rejected.

Object birth is not creation of new underlying states.

It is creation of new effective distinctions.

The states were already present.

The ontology changes because the variational balance or admissibility conditions change.

This provides a precise mathematical interpretation of conceptual discovery.

Discovery can occur as refinement of the optimizing equivalence relation.

A new object is born when the cost of treating its states as part of an older class becomes too high.

9.5 Plateaus of Stable Ontology

Between critical values, the minimizer family remains constant.

If one policy $E^\star$ is uniquely optimal throughout an interval

$\lambda \in (\lambda a, \lambda b)$,

then the quotient ontology

$X/E^\star$

forms a stable plateau of conceptual resolution.

Changes in λ inside that interval do not change the selected objects.

The objective values change because complexity is reweighted, but no competing policy becomes preferable.

The width

$\lambda b - \lambda a$

is one measure of ontological robustness.

A wide plateau indicates that the policy remains optimal under substantial variation in the price of complexity.

A narrow plateau indicates fragility.

The policy is selected only under a finely tuned tradeoff.

Plateau width is not the only robustness measure.

The objective gap at a specific λ also matters.

Let

$\gamma(\lambda) =$
minimum cost of a nonoptimal policy

minus
the optimal cost.

A large γ means that substantial perturbation is required before another ontology becomes competitive.

A small γ means that the system is near a transition.

One may also vary several parameters simultaneously.

Let θ represent a vector containing:

λ,

distortion weights,

observer reliabilities,

invariance penalties,

time horizons,

and other task parameters.

The region

R(E) =
{θ : E is optimal under θ}

is the stability region of policy E.

In one parameter, R(E) is an interval or union of intervals.

In several parameters, it becomes a cell or region in parameter space.

The boundaries between regions are transition surfaces.

A robust object system occupies a large region.

A fragile system occupies a small or thin region.

Individual objects can also be assigned persistence measures.

An equivalence pair (x, y) is persistent when it belongs to the optimal relation across a broad parameter range.

A class C is persistent when substantially the same group of states remains unified across many optimal policies.

A distinction is persistent when two states remain separate across the parameter region.

This suggests a graded notion of ontological confidence.

An identity accepted only under one narrow weighting is weak.

An identity retained across many tasks and parameter regimes is strong.

A robust object can therefore be defined as a class or identity structure stable across a family of variational problems.

This is stronger than optimality at one λ.

It approaches task-independent objecthood without assuming absolute context independence.

Stable plateaus also provide a practical method for selecting λ.

Rather than choose one value arbitrarily, compute the resolution path and identify broad plateaus.

A policy persisting across a wide interval may be preferred over one appearing only at a transition point.

This resembles scale selection in clustering and model selection.

Stable resolutions are more credible than isolated optima produced by precise tuning.

9.6 Bifurcations Among Optimal Policies

At a critical value, the minimizer set may contain more than one policy.

This is an ontological bifurcation.

Two or more object systems are equally favored under the current variational law.

The simplest case contains two policies E and F.

They may be nested:

$E \preceq F$.

Then the bifurcation represents a direct choice between fine and coarse resolution.

They may also be incomparable.

In that case, the system faces two different ways of organizing the state space.

Each policy preserves distinctions the other erases.

The bifurcation is not merely about how many objects exist.

It concerns which differences define the objects.

Incomparable bifurcation can occur when two alternative explanatory or predictive structures have equal cost.

One ontology groups states by one feature.

Another groups them by a different feature.

Both achieve the same distortion-complexity balance.

Such degeneracy may arise from symmetry.

Suppose the state space and objective are invariant under exchanging two coordinates.

A policy grouping by the first coordinate and a policy grouping by the second may tie.

The bifurcation reflects symmetry of the problem.

Degeneracy can also reflect insufficient data.

Two ontologies may make indistinguishable predictions on available observations.

Further experiments are required to decide between them.

A bifurcation can therefore guide active observation.

The system should seek a measurement or intervention whose outcomes differ under the competing policies.

That observation breaks the degeneracy by altering $D\pi$.

A third possibility is to retain multiple object systems.

If different minimizers are useful for different subcontexts, forcing one global choice may be unnecessary.

The ontology can remain plural.

A higher-level structure may index the alternative quotients by task or policy.

Bifurcation points are also sensitive to perturbation.

A small change in data, weighting, or prior policy may select one branch.

This is symmetry breaking in the optimization landscape.

The chosen ontology may then depend on history, initialization, or noise even when the unperturbed problem is symmetric.

One should therefore distinguish:

mathematical degeneracy,

algorithmic selection,

and scientific justification.

An optimization algorithm returning one branch does not prove that branch is uniquely warranted.

The complete minimizer set must be considered.

A bifurcation diagram can represent optimal policies as a function of λ or another control parameter.

Branches correspond to policy families.

Crossings correspond to equal objective value.

Stable branches lie on the lower envelope.

Metastable branches may represent local minima that are not globally optimal.

This diagrammatic approach provides a global view of conceptual alternatives.

9.7 Hysteresis and Policy Dependence

The basic objective

$D\pi(E) + \lambda K\pi(E)$

depends only on the current policy and current parameter values.

Under exact global optimization, the selected minimizer at a given λ does not depend on the path used to reach that value, except when several minimizers tie.

Real conceptual systems often exhibit history dependence.

An established ontology has inertia.

Changing labels, models, databases, institutional practices, scientific vocabulary, or learned representations has cost.

To represent this, include a revision term relative to the previous policy Eold:

Jhist(E; Eold) =
Dπ(E)
+

λKπ(E)
+
τdEq(E, Eold).

The parameter τ ≥ 0 is the cost of ontological change.

The function dEq measures the difference between policies.

It may count changed pairwise identity judgments, class mergers and splits, or reclassification cost.

Because the objective now depends on Eold, the selected policy can depend on history.

Suppose λ increases gradually.

The current fine policy may remain preferred beyond the transition point of the memoryless objective because changing ontology carries cost.

Eventually the complexity benefit becomes large enough to justify fusion.

If λ later decreases, the coarse policy may persist below the original transition because splitting it also carries revision cost.

The upward and downward transition values differ.

This is hysteresis.

Hysteresis produces a loop in the relation between λ and an order parameter such as object count.

The ontology remembers its path.

This phenomenon is not necessarily irrational.

Conceptual change has real cost.

Scientific communities must revise terminology and retrain practitioners.

AI systems must reorganize memory and downstream representations.

Databases must migrate entity links.

Organisms and institutions may be unable to revise classifications instantly.

However, hysteresis can also preserve obsolete objects.

A policy may remain in use after another has become intrinsically better because transition cost is high.

The variational framework separates current adequacy from revision burden.

Policy dependence also arises when the distortion function itself is learned through the current ontology.

A representation determines which evidence is noticed.

The evidence then reinforces the representation.

This creates feedback.

Let $D_\pi(E; E_{old})$ depend on features or predictions learned under E_{old}.

The optimization may become path-dependent even without an explicit revision term.

Different initial policies can lead to different local optima.

Such dependence is especially important in artificial intelligence.

An AI that learns identity policies incrementally may become trapped in a conceptual basin.

A mistaken early merger can alter future data aggregation.

A mistaken early split can prevent recognition of a broader pattern.

Corrective mechanisms therefore require retained provenance, alternative hypotheses, and periodic global reoptimization.

Hysteresis also distinguishes reversible and irreversible object formation.

A reversible transition returns to the original policy when the control parameter is restored.

An irreversible transition does not, because information was discarded, training changed, or the feasible region itself evolved.

Coarsening is particularly prone to irreversibility.

Once internal distinctions are forgotten, later refinement may be impossible without access to the original states.

Thus the mathematical asymmetry between quotienting and refinement becomes a practical asymmetry in ontology dynamics.

9.8 Emergence as an Optimization Phase Transition

Emergence is often described informally as the appearance of a higher-level object not evident in the individual components.

The variational theory provides a precise version of this idea.

An emergent object appears when a coarser equivalence class becomes part of the optimal identity policy under a change in scale, task, parameter, or resource constraint.

Let E be the earlier policy.

Let F be the later policy.

Suppose $E \preceq F$ and several E-classes merge into one F-class C.

The class C is emergent relative to E when:

it was not an object under the finer policy,

it becomes one object under the optimizing policy F,

its merger yields sufficient compression,

its task distortion remains acceptable,

and it satisfies the required descent and transformation conditions.

The object is not added to X.

It is formed by a newly optimal rule of identity.

This explains how an emergent object can be real at one level while absent at another.

At the detailed level, the component states remain distinct.

At the effective level, their differences do not justify separate object status.

The macro-object is the quotient class selected by the variational law.

Transformation invariance is essential.

A merely compressed cluster is not necessarily an object.

The class must support sufficiently autonomous behavior.

Descent is also essential in distributed systems.

The class must not destroy protected local distinctions unless their revision is explicitly accepted.

Emergence therefore requires more than aggregation.

It requires variationally favorable, dynamically stable, and observationally coherent aggregation.

Object birth through refinement is another form of emergence.

A new subtype, phase, or causal category becomes effective when internal differences acquire enough task value to justify additional complexity.

Thus emergence includes both fusion and differentiation.

The direction depends on which change in policy becomes optimal.

Several kinds of emergence can be distinguished.

Compression-driven emergence occurs when λ increases and broader objects become economical.

Evidence-driven emergence occurs when new data change distortion and reveal a new stable class.

Dynamics-driven emergence occurs when a quotient acquires autonomous behavior at a particular time scale.

Descent-driven emergence occurs when several local object traces form one coherent global class.

Intervention-driven emergence occurs when a common response profile defines a new causal object.

Symmetry-driven emergence occurs when many states become equivalent under an admissible transformation group.

These mechanisms share one form:

the optimizer moves to a new region of Eq(X).

The phase-transition framework also clarifies levels of reality.

An object stable only at one finely tuned λ is weakly effective.

An object persisting across a wide parameter plateau is robust.

An object selected across many tasks and transformation families is more deeply invariant.

A hierarchy of objecthood can therefore be based on the measure of the parameter and policy region in which an identity remains optimal.

One may define an object persistence score

P(C) =
measure of the parameter region in which C appears as a class or stable trace of an optimal policy.

The precise definition requires methods for matching classes across policies.

A high persistence score indicates robust emergence.

A low score indicates fragile or task-specific objecthood.

The phase theory therefore provides a bridge between policy-relative and robust objecthood.

Objects remain relative to what must be preserved and what complexity costs.

Yet some objects appear across broad ranges of those choices.

They are not absolute by assumption.

They are stable extremizers.

The principal conclusions of this section are:

Feasible identity policies define a compression-distortion frontier.

Weighted optimization selects supported frontier policies.

Critical parameter values occur when competing policy costs become equal.

Object fusion occurs when a coarser policy becomes optimal.

Object birth occurs when refinement becomes worth its added complexity or is required by new evidence or constraints.

Intervals of unchanged optimal policy form stable ontological plateaus.

Degenerate minima create bifurcations among alternative object systems.

Revision costs produce hysteresis and historical dependence.

Emergence can be represented as a phase transition in the optimizing equivalence relation.

The next section turns from analysis to discovery algorithms. It develops methods for learning identity policies, searching the partition lattice, evaluating merge and split operations, using differentiable relaxations, and designing active observations that resolve ontological uncertainty.

10. Learning and Discovering Identity Policies

The variational theory becomes operational when a system can search for identity policies rather than merely evaluate policies supplied in advance.

This is the computational transition from classification to object discovery.

Ordinary supervised learning begins with labels. The learner estimates a map from observations into an already defined object space.

Ordinary clustering begins with a similarity measure, a distance, a model family, or a prescribed number of clusters. The learner groups states according to criteria chosen before the optimization begins.

Variational object discovery is more demanding.

The learner must determine:

which differences are irrelevant,

which substitutions are lawful,

how many quotient objects should exist,

which transformations those objects must survive,

which local distinctions must remain protected,

and how much complexity reduction justifies a given loss of fidelity.

The output is not merely a classifier.

It is an equivalence relation $E^\star$ and its quotient $X/E^\star$.

The system is therefore learning the ontology in which later classification, prediction, explanation, and intervention will occur.

This task is combinatorial. For a finite state space with n elements, the number of candidate identity policies is the nth Bell number. Exhaustive search becomes infeasible quickly.

The lattice structure of Eq(X) must therefore guide the computation.

Local policy changes can be expressed as class mergers and class splits.

Closure operators can project candidate policies into coherent regions.

Constraint tests can reject transformations producing descent or invariance defects.

Approximate relaxations can replace hard partitions with soft assignments during optimization.

Active observation can identify which new measurements would most efficiently resolve uncertainty among competing policies.

The computational problem also contains an unavoidable epistemic difficulty.

The distortion functional is usually not known perfectly. It must be estimated from finite data, learned models, observed transformations, interventions, or explanatory performance. An identity policy may appear lawful under current evidence and fail after new observations.

Object discovery must therefore remain revisable.

A robust system should retain the provenance of equivalence decisions, competing candidate policies, and the evidence that would cause classes to merge or split.

This section develops the main computational strategies for variational objecthood. It distinguishes fixed-label learning from identity-policy learning, formulates search over partitions, defines local merge and split operations, examines non-gradient methods and differentiable relaxations, introduces constraint-guided clustering, and presents active observation as a method for discovering the correct conceptual resolution.

10.1 From Fixed Labels to Learned Equivalence

In supervised classification, the training data typically contain pairs

(x, y),

where x is an observed state and y is a supplied class label.

The learner estimates a function

f: X → Y.

The induced identity policy is

x E y

exactly when

f(x) = f(y).

The ontology Y is supplied by the dataset.

The learner does not decide whether two labels should merge, whether one label should split, or whether an entirely different partition would better preserve the relevant transformations and outcomes.

The labels may themselves encode historical, institutional, scientific, or annotator assumptions.

A system trained to reproduce them inherits those assumptions.

Variational object discovery reverses the direction.

The states, tasks, transformations, and constraints are supplied.

The equivalence relation is learned.

The learner searches for E satisfying

$E^{\star} \in \arg\min \{D\pi(E) + \lambda K\pi(E) : E \in A\pi\}$.

The quotient labels are outputs, not inputs.

Even when partial labels are available, they need not determine the final ontology.

They can enter as:

must-link constraints,

cannot-link constraints,

soft similarity evidence,

or protected local distinctions.

The system may discover that several supplied labels are behaviorally equivalent and should merge.

It may discover that one label contains states with different predictions or interventions and should split.

It may also find that the available evidence does not determine one unique policy.

Pairwise supervision provides one natural input format.

A must-link pair

(x, y)

asserts that x and y should be equivalent.

A cannot-link pair asserts that they should remain distinct.

Must-link declarations generate equivalence closure.

Cannot-link declarations constrain which generated components are admissible.

The conflict between the two is exactly a gluing-style consistency problem.

If must-link closure connects a cannot-link pair, no equivalence relation can satisfy all declarations.

The system must revise or weight the constraints.

Transformation supervision provides a stronger source of information.

Suppose observations show that x and y produce the same task-relevant outcomes under every tested transformation. This supports their merger.

If one transformation separates their consequences, the distinction becomes relevant.

Thus equivalence can be learned from behavioral interchangeability rather than from externally assigned names.

A learned identity policy should be distinguished from a learned embedding.

An embedding

$z: X \to \mathbb{R}d$

places states in a representation space.

It does not by itself define lawful equivalence.

One must still decide which points should be treated as one object and verify that this decision is stable under the relevant operations.

The equivalence relation may be generated by:

equal embeddings,

thresholded distances,

shared predictions,

orbit membership,

or another learned criterion.

The relation must then be checked for transitivity, descent, and transformation invariance.

Learning a useful geometry is not the same as learning objecthood.

The latter requires a policy of admissible substitution.

10.2 Search over Partitions

For finite X, direct object discovery is a combinatorial optimization over partitions.

A partition can be represented in several ways.

One representation stores a class label

$c(x) \in \{1, \ldots, k\}$

for each state.

Two states are equivalent when their labels agree.

The numerical labels have no intrinsic meaning. Permuting them produces the same partition.

Algorithms must therefore avoid treating label permutations as different policies.

A second representation stores the classes directly:

E = {C1, …, Ck}.

A third representation uses a binary matrix M with entries

Mxy =
1 if x E y,

0 otherwise.

The matrix must be reflexive, symmetric, and transitive.

A fourth representation uses a graph of generating identity edges. The equivalence classes are its connected components.

The graph representation is sparse but redundant. Many edge sets generate the same partition.

Each representation supports different computational operations.

Label vectors are convenient for local reassignment.

Class lists are convenient for merge and split moves.

Pairwise matrices are convenient for constraint programming.

Generator graphs are convenient for provenance and distributed identity.

The search can begin from either extreme.

An agglomerative procedure starts with equality:

E0 = ΔX.

Every state is initially its own object.

The algorithm repeatedly merges classes when the merger reduces the total objective and preserves all hard constraints.

This approach begins with zero state-merging distortion and gradually increases abstraction.

A divisive procedure starts with universal equivalence:

E0 = ∇X.

All states begin as one object.

The algorithm repeatedly splits classes when the reduction in distortion or defect justifies the added complexity.

This approach begins with maximum compression and repairs destructive fusion.

Neither direction is universally preferable.

Agglomerative search may become trapped because an early merger prevents a later better organization.

Divisive search may make poor initial splits that require reassembly.

Bidirectional search allows both operations.

The search graph of Eq(X) can be defined by elementary cover relations in the partition lattice.

A policy F covers E upward when F is obtained by merging exactly two E-classes.

A policy E covers F downward when E is obtained by splitting one F-class into exactly two blocks.

More complex merges and splits can be decomposed into these elementary operations.

The objective landscape can then be explored through local neighborhoods.

A basic local-search algorithm is:

initialize E,

enumerate admissible neighboring merges and splits,

evaluate the objective change,

choose an improving move,

update E,

repeat until no improving move remains.

The procedure terminates on finite X if every accepted move strictly reduces the objective.

It may terminate at a local minimum.

Global methods include:

branch and bound,

integer programming,

dynamic programming under special structure,

simulated annealing,

tabu search,

evolutionary methods,

Markov chain sampling over partitions,

and multi-start local search.

The feasibility constraints can reduce the search space substantially.

Transformation invariance restricts candidate partitions to Inv(G).

Exact observable preservation restricts E to a subrelation of common kernels.

Protected distinctions prohibit selected merges.

Closure coherence restricts distributed families to Fix($C\mathcal{U}$).

Where possible, the search should operate directly inside these constrained lattices rather than repeatedly proposing infeasible policies.

10.3 Local Merge and Split Operators

Merge and split operations are the elementary moves of conceptual revision.

Let E have quotient classes

C1, C2, …, Ck.

A merge operator selects two or more classes and replaces them with their union.

For a binary merge of A and B, define

MergeE(A, B)

as the equivalence relation whose classes are:

A ∪ B,

and every original E-class other than A and B.

The complexity change is usually favorable:

K(MergeE(A, B)) ≤ K(E).

The distortion change is usually unfavorable:

Dπ(MergeE(A, B)) ≥ Dπ(E)

under monotone within-class distortion.

Define the merge score

$$S_{merge}(A, B \mid E) = J_{\pi,\lambda}(Merge_E(A, B)) - J_{\pi,\lambda}(E).$$

A negative score means the merger improves the objective.

Hard constraints must be checked before acceptance.

The merger may violate:

protected separation,

transformation invariance,

descent coherence,

or exact observable preservation.

A split operator selects one class C and partitions it into nonempty blocks

$C_1, \ldots, C_m$.

The resulting policy is finer.

The complexity normally increases.

The distortion normally decreases.

Define

$$S_{split}(\{C_1, \ldots, C_m\} \mid E) = J_{\pi,\lambda}(Split_E(C; C_1, \ldots, C_m)) - J_{\pi,\lambda}(E).$$

A negative score means the split is favorable.

The difficult part is proposing useful splits.

The number of possible partitions of one large class is itself enormous.

Several heuristics are available.

Predictive splitting separates states according to different outcomes or predictive distributions.

Dynamical splitting separates states according to different successor classes.

Intervention splitting separates states according to different response profiles.

Descent splitting separates states involved in different local object traces.

Geometry-based splitting uses a learned metric or embedding.

Explanation-based splitting separates states requiring different models.

Defect-driven splitting is especially important.

If a class C fails transformation invariance because g(C) intersects several quotient classes, split C according to the image class under g.

If C contains a protected pair, split it so that the pair is separated.

If a global class causes a local descent collision, split one or more generating local classes along the obstruction witness.

These splits are not arbitrary. They are certified by structural failure.

Merge proposals can also be structured.

Only classes with similar predictions, transformations, or explanations need be considered.

Cross-observer correspondences may suggest mergers of local object traces.

Symmetry orbits may define natural merge candidates.

A merger should ideally be evaluated through its full closure consequences.

In distributed systems, merging A and B can connect larger components and create remote local collisions.

The marginal cost is therefore nonlocal.

Efficient incremental data structures are needed to update:

connected components,

class statistics,

distortion summaries,

and constraint violations.

The merge-split process resembles scientific concept revision.

A merge asserts that a distinction is unnecessary.

A split asserts that a previously ignored difference matters.

The variational score gives each revision a precise justification.

10.4 Gradient-Free Policy Optimization

The partition lattice is discrete.

Ordinary gradient descent does not apply directly to hard equivalence relations.

Gradient-free methods are therefore natural.

Greedy agglomeration selects the best currently available merge.

At each step, compute

$\Delta J(A, B) =$
$J(\text{MergeE}(A, B)) - J(E).$

Choose the admissible pair with the most negative ΔJ.

Stop when no merger improves the objective.

This method is simple and interpretable.

Its result depends on early choices.

A locally beneficial merge can block a later globally better partition.

Greedy division applies the same idea to splits.

A combined hill-climbing method allows both.

Simulated annealing occasionally accepts worse moves according to a temperature parameter. This can escape local minima.

The probability of accepting an objective increase ΔJ may be proportional to

$\exp(-\Delta J/T).$

As T decreases, the search becomes more selective.

Tabu search prevents immediate reversal of recent moves and encourages exploration of new policy regions.

Evolutionary methods represent partitions as individuals and use mutation through merges, splits, or state reassignment.

Crossover between partitions is difficult because arbitrary combinations may violate equivalence structure or constraints.

Markov chain methods define a probability distribution over policies:

P(E) proportional to exp(−βJ(E)).

Sampling reveals not only the global optimum but also near-optimal ontologies and uncertainty over policies.

At large β, the distribution concentrates near minimizers.

This provides a statistical-mechanical interpretation of the objecthood landscape.

Beam search retains several best policies at each resolution level.

It can preserve alternative ontological branches that a greedy algorithm would discard.

Branch and bound uses lower bounds on the best possible objective within a subtree of partition decisions.

If the bound exceeds the current best solution, that subtree can be pruned.

Exact optimization may be possible for modest finite problems.

Constraint programming can encode equivalence variables and transitivity conditions.

For binary pair variables exy, transitivity can be enforced through inequalities such as

exy + eyz − 1 ≤ exz.

Additional constraints encode must-links, cannot-links, and transformation preservation.

The objective can include pairwise distortion and complexity approximations.

The resulting integer program may be difficult but exact when solved.

The best method depends on the structure of Dπ and Kπ.

If the objective decomposes over classes, local updates are efficient.

If transformations or descent create nonlocal dependencies, global recomputation may be necessary.

Gradient-free optimization has one major advantage.

Every candidate is a genuine identity policy.

There is no ambiguity about whether the result satisfies transitivity.

Its main disadvantage is combinatorial cost.

Differentiable relaxation provides an alternative.

10.5 Relaxed Differentiable Approximations

A hard equivalence relation can be approximated by a soft assignment.

Suppose a maximum of k latent objects is allowed.

Assign to each state x a probability vector

$p(x) = (p_1(x), \ldots, p_k(x))$,

where

$p_j(x) \geq 0$

and

$\sum_j p_j(x) = 1$.

The value $p_j(x)$ represents the degree or probability with which x belongs to latent object j.

A hard partition is recovered when each vector is one-hot.

Soft assignments permit differentiation with respect to model parameters.

A soft same-object score can be defined by

$s(x, y) = \sum_j p_j(x) p_j(y)$.

This quantity is large when x and y have similar membership distributions.

It is symmetric and reflexive only approximately.

It does not automatically satisfy transitivity in the hard logical sense.

The relaxation must therefore include regularization encouraging crisp and coherent assignments.

An entropy penalty

$$H_{\text{assign}} = -\sum_x \sum_j p_j(x) \log p_j(x)$$

can encourage or discourage sharp assignments depending on its sign.

A balance penalty can prevent collapse into one cluster when such collapse is not supported by the task.

Transformation consistency can be imposed by requiring

$p(g(x))$

to be predictable from p(x) through a quotient-level map.

Descent consistency can compare local soft assignments on overlaps.

Protected distinctions can penalize large s(x, y) for cannot-link pairs.

Must-link constraints can penalize differences between p(x) and p(y).

A differentiable objecthood loss may take the form

L =
Ltask
+
λLcomplexity
+
μLdescent
+
νLinvariance
+
ρLseparation
+
ηLdiscreteness.

The learned soft representation is then projected to a hard partition.

Projection may assign each state to its largest membership component.

More sophisticated rounding can optimize the hard objective near the relaxed solution.

The relaxation creates several challenges.

First, the number k of latent objects may need to be bounded in advance.

One can use an oversized k and allow unused components, but this introduces identifiability and regularization issues.

Second, soft similarity is not exact equivalence.

The projected partition may have worse objective value or violate constraints.

Third, label symmetry creates many equivalent parameter settings.

Fourth, the relaxed loss may favor diffuse assignments that do not correspond to meaningful object classes.

Fifth, local minima remain possible.

Despite these issues, differentiable methods are important for large state spaces and learned representations.

The state set may be implicit or continuous.

The equivalence policy may be generated by a neural encoder and assignment network rather than stored explicitly.

One may learn a function

$z_\theta: X \to Z$

and define equivalence through a learned quantizer or invariant code.

The transformation, prediction, and intervention losses train the representation so that states assigned to one code behave similarly.

The variational theory supplies the missing structural tests.

After training, the induced hard policy should be checked for:

task distortion,

transformation invariance,

descent coherence,

protected separation,

and stability across parameter changes.

A low neural loss is not by itself a proof of lawful objecthood.

10.6 Constraint-Guided Clustering

Variational object discovery can be viewed as clustering under stronger structural constraints than ordinary clustering.

The data may supply a similarity or dissimilarity measure

$d(x, y)$.

A conventional clustering method groups states with small within-class distances.

Constraint-guided clustering adds:

must-link relations,

cannot-link relations,

transformation compatibility,

local-to-global descent,

and task-relevant distortion.

The objective may be written schematically as

J(E) =
within-class task loss
+
λ quotient complexity
+
constraint penalties.

The central difference is that geometric proximity is only evidence.

It does not automatically license identity.

Two nearby states may respond differently to intervention.

Two distant states may be equivalent under symmetry.

The correct class geometry is induced by lawful substitution, not by raw distance alone.

Must-link constraints can be closed transitively before clustering.

This produces required components.

Cannot-link constraints between states in the same required component reveal infeasibility immediately.

At the component level, the clustering problem becomes one of merging must-link components while respecting cannot-links and other structural laws.

Transformation constraints can be represented as implications.

If x E y, then g(x) E g(y).

In pair-variable form,

exy ≤ eg(x)g(y).

For several transformations, these constraints propagate identity decisions through the transformation graph.

Protected separation is

exy = 0

for selected pairs.

Required identity is

exy = 1.

Descent constraints can be represented by prohibiting one global component from containing two distinct protected local classes from the same domain.

Constraint-guided clustering can therefore be implemented on the local-object graph.

Nodes represent locally formed classes.

Edges represent candidate global identity.

Some edges are required.

Some are forbidden.

Some carry costs or confidence.

The goal is to choose a globally coherent set of mergers minimizing the objecthood functional.

This is closely related to correlation clustering, constrained partitioning, and entity-resolution optimization, but with explicit closure and transformation structure.

The output should include an obstruction certificate whenever constraints are inconsistent.

A system should not silently violate a cannot-link because enough must-link evidence points the other way.

It should report the identity path causing the conflict.

Constraint guidance also improves interpretability.

Every merger can be explained by the observations, transformations, or compression gain supporting it.

Every split can be explained by distortion or defect.

The resulting ontology is auditable.

10.7 Active Observation for Identity Discovery

When several policies remain feasible or nearly optimal, the system can seek new evidence that distinguishes them.

This is active observation.

Let

M = {E1, …, Em}

be a set of competing identity policies.

A candidate observation a may be:

a measurement,

an intervention,

a new sensor placement,

an added overlap domain,

a longer time horizon,

or a targeted comparison between states.

The observation produces a possible result r.

Each candidate policy predicts a distribution

P(r | a, E).

An informative observation is one whose possible outcomes sharply reduce uncertainty over the policy.

One criterion is expected minimizer-set reduction.

Choose a maximizing

expected decrease in the number of viable policies.

Another criterion is expected information gain:

IG(a) =
H(E | current data)
–
expected H(E | current data and result of a).

A decision-focused criterion chooses the observation most likely to change the selected quotient or reduce uncertainty in task value.

A descent-focused observation targets underdetermined global identities.

Suppose two global components could be merged without contradicting existing local policies.

Add a domain containing representative states from both components.

A local measurement can determine whether they should remain distinct.

An obstruction-focused observation targets a defective identity path.

Suppose a global class fuses two protected local objects through several uncertain bridge edges.

Measure the weakest or most consequential bridge.

If the bridge is rejected, the false fusion may dissolve.

A transformation-focused observation applies an action to two states currently treated as equivalent.

If their responses differ, the class should split or the task policy should be revised.

An explanation-focused observation tests competing mechanisms assigned to one quotient object.

Active observation therefore closes the loop between ontology and experiment.

The current identity policy determines what uncertainty matters.

The selected experiment generates evidence.

The evidence changes the distortion functional, constraints, or transformation model.

The identity policy is reoptimized.

This produces an iterative discovery cycle:

form provisional objects,

identify uncertainty or defect,

select a discriminating observation,

update evidence,

revise the policy,

and repeat.

The system does not passively wait for arbitrary data.

It seeks observations capable of resolving objecthood.

This is especially important because many identity questions cannot be settled from passive similarity alone.

Intervention reveals causal distinction.

Motion reveals persistence.

New overlaps reveal global identity.

Perturbation reveals hidden instability.

Active observation converts conceptual uncertainty into experimental design.

10.8 Artificial Intelligence as Self-Revision of Object Space

An artificial intelligence system usually operates inside a supplied representational ontology.

Its tokens, labels, latent dimensions, database entities, and memory units are designed or trained under fixed architectural assumptions.

Variational objecthood suggests a stronger form of intelligence.

The system should be able to revise what counts as one object.

Such an AI would maintain:

a state or experience space X,

a current identity policy E,

a quotient memory or concept space X/E,

a family of admissible transformations G,

task-relevant distortion estimates,

complexity costs,

and local or distributed coherence constraints.

It would monitor the adequacy of its current ontology.

A concept should split when states inside it develop incompatible predictions, interventions, transformations, or explanations.

Concepts should merge when their distinction no longer contributes enough value to justify its complexity.

A local identity should be revised when global closure produces a certified defect.

A new observation should be requested when several policies remain nearly optimal.

The update process can be written:

$$
\begin{aligned}
&E_{t+1} \in \arg\min \{ \\
&D_{\pi t}(E) \\
&+ \\
&\lambda_t K_{\pi t}(E) \\
&+ \\
&\mu_t L_{des}(E) \\
&+ \\
&\nu_t S_{inv}(E) \\
&+ \\
&\tau_t d_{Eq}(E, E_t) \\
&: \\
&E \in E_q(X_t) \\
&\}.
\end{aligned}
$$

The term $d_{Eq}(E, E_t)$ penalizes abrupt ontology change.

The state space X_t may also grow as new experiences arrive.

The task policy π_t and cost parameters may change over time.

This equation describes ontology learning rather than ordinary parameter learning.

The AI is not only updating beliefs about objects.

It is updating the equivalence relation that determines the objects themselves.

Several safeguards are necessary.

First, provenance must be retained.

The system should know which observations and transformations support each merger.

Second, protected distinctions must be explicit.

Safety-critical categories should not disappear merely because compression is rewarded.

Third, local and global alternatives should be retained when the optimum is nonunique.

Fourth, revision should be reversible when possible.

Underlying states should not be discarded merely because the current quotient ignores their differences.

Fifth, the system should detect objective misspecification.

A quotient can be optimal under the wrong distortion function.

Sixth, human or institutional labels should be treated as policies and evidence, not automatically as ground truth.

Seventh, the AI should distinguish similarity, correlation, and identity.

Only the last licenses general substitution.

An ontology-learning AI would have several advantages.

It could adapt its conceptual resolution to new tasks.

It could discover macrostates and invariants.

It could integrate distributed observations without naive transitive fusion.

It could explain why concepts merged or split.

It could design experiments to resolve uncertainty.

It could compress memory while protecting predictive and causal structure.

It could maintain several policy-indexed ontologies rather than forcing every task into one universal classification.

This last point is important.

The optimal identity policy may depend on π.

An intelligent system should not necessarily seek one final ontology.

It may maintain a family

$\{E\pi : \pi \in \Pi\}$

indexed by task, observer, scale, or admissibility regime.

Relations among these policies can be organized by refinement and translation.

A causal ontology may refine a predictive ontology.

A safety ontology may refine a convenience ontology.

A long-horizon ontology may differ from a short-horizon ontology.

The system's intelligence lies partly in selecting the correct object space for the current problem.

The computational program of variational objecthood can therefore be summarized.

Represent identity policies explicitly.

Evaluate them by task distortion and effective complexity.

Enforce or penalize descent and transformation defects.

Search through merge and split operations.

Use relaxations when direct partition search is too large.

Retain alternative near-optimal policies.

Actively acquire evidence that resolves identity uncertainty.

Revise the quotient ontology over time.

This is a path toward artificial intelligence that discovers its own conceptual units.

The system would not merely classify a predivided world.

It would learn which divisions are lawful, economical, stable, and useful.

The next section applies these methods to explicit model constructions. It develops finite classification examples, dynamical state lumping, multisensor object formation, image invariance, scientific taxonomy, distributed identity penalties, emergent physical macrostates, and computational searches for optimal quotients.

11. Examples and Model Constructions

The preceding theory is abstract because it is intended to apply across mathematics, physics, biology, artificial intelligence, distributed sensing, and scientific classification. Its central variable is always an identity policy, but the meaning of distortion, complexity, invariance, and descent changes with the application.

This section develops explicit constructions showing how the framework operates.

The examples are deliberately finite or structurally simplified so that every equivalence relation, quotient map, constraint, and objective term can be inspected. They are proofs of mathematical structure rather than substitutes for full empirical implementations.

The examples demonstrate eight recurring patterns.

First, a classification problem may contain both task labels and protected distinctions.

Second, the same dynamical system may admit several invariant quotients, with the task selecting among them.

Third, multisensor identity can fail through a global path even when no pair of observers directly contradicts another.

Fourth, scientific taxonomies depend on which traits, mechanisms, or interventions must be preserved.

Fifth, image objects can be formed by quotienting nuisance transformations while protecting semantic distinctions.

Sixth, distributed identity can be optimized by balancing correspondence confidence against descent penalties.

Seventh, physical macrostates can be interpreted as compressed quotient objects whose adequacy depends on autonomous dynamics.

Eighth, finite computational searches can discover optimal policies and report exact certificates of their selection.

These constructions show that the variational objecthood principle is not one particular clustering algorithm. It is a general architecture within which many kinds of object formation can be represented, constrained, optimized, and audited.

11.1 Finite Classification with Protected Distinctions

Consider the finite state space

$X = \{a, b, c, d, e, f\}$.

Suppose a task-relevant observable h assigns three outcome types:

$h(a) = 0$,

$h(b) = 0$,

$h(c) = 1$,

$h(d) = 1$,

$h(e) = 2$,

$h(f) = 2$.

If outcome prediction were the only requirement, the coarsest zero-distortion policy would be

Eh =
{{a, b}, {c, d}, {e, f}}.

Each class contains states with one common h-value.

The quotient has three objects.

Now impose one protected distinction:

a must remain distinct from b.

This distinction may represent a safety condition, a causal difference not expressed by h, a legal separation, or an experimentally established disequality.

The relation Eh is no longer feasible because it identifies a and b.

The coarsest zero-distortion policy satisfying the protected distinction is

E0 =
{{a}, {b}, {c, d}, {e, f}}.

Its quotient has four objects.

Let complexity be quotient cardinality:

K(E) = |X/E|.

Let predictive distortion be the fraction of states misclassified by the best single h-value assigned to each quotient class.

For E0,

D(E0) = 0,

K(E0) = 4.

Now consider the coarser policy

E1 =
{{a}, {b, c, d}, {e, f}}.

The protected pair remains separate because a is not equivalent to b.

The class {b, c, d} contains one state with label 0 and two states with label 1. Its best common prediction is 1, producing one error.

Thus

D(E1) = 1/6,

K(E1) = 3.

Consider a still coarser feasible policy

E2 =
{{a}, {b, c, d, e, f}}.

The large class contains labels

0, 1, 1, 2, 2.

The best common prediction can be either 1 or 2, producing three errors among the six total states.

Therefore

D(E2) = 1/2,

K(E2) = 2.

The universal relation is infeasible because it would merge the protected pair a and b.

For the three candidate policies, the objecthood costs are

Jλ(E0) = 4λ,

Jλ(E1) = 1/6 + 3λ,

Jλ(E2) = 1/2 + 2λ.

The first transition occurs when E0 and E1 have equal cost:

4λ = 1/6 + 3λ.

Therefore

λ = 1/6.

Below λ = 1/6, E0 is preferred among these policies.

The exact prediction supplied by four objects is worth its additional complexity.

Above λ = 1/6, E1 becomes preferable to E0.

The system accepts one prediction error in exchange for reducing the quotient from four objects to three.

The second transition occurs when E1 and E2 tie:

1/6 + 3λ = 1/2 + 2λ.

Therefore

$\lambda = 1/3$.

Above $\lambda = 1/3$, E2 is preferred among the three.

The system compresses five states into one broad object while preserving the mandatory separation between a and b.

This example displays several features of the general theory.

The task observable alone does not determine the feasible quotient.

The protected distinction forces additional refinement.

Zero distortion does not imply equality-level resolution.

The policy E0 compresses six states into four objects while preserving h exactly.

Increasing the cost of complexity produces discrete transitions.

The selected ontology changes from four objects to three and then to two.

The hard constraint prevents complete collapse.

The example also distinguishes task equivalence from total identity.

The observable h regards a and b as interchangeable.

The protected-separation policy does not.

Thus the effective ontology preserves more structure than the prediction task alone requires.

This is common in real applications.

A medical prediction model may assign two patients the same risk while treatment constraints require them to remain distinct.

A database may assign two records the same category while legal identity requires separate entities.

An AI system may find two states behaviorally similar while a safety policy prohibits substitution.

The feasible identity policy is therefore determined by the intersection of task preservation and admissibility constraints.

11.2 Dynamical State Lumping

Consider a deterministic dynamical system with state space

$X = \{1, 2, 3, 4\}$

and transformation T defined by

$T(1) = 2,$

$T(2) = 1,$

$T(3) = 4,$

$T(4) = 3.$

The system consists of two independent two-cycles:

$1 \leftrightarrow 2,$

$3 \leftrightarrow 4.$

Several nontrivial invariant equivalence relations exist.

Define

EA =
{{1, 3}, {2, 4}}.

Under T, the first class maps to the second:

$T(\{1, 3\}) = \{2, 4\}.$

The second maps to the first:

$T(\{2, 4\}) = \{1, 3\}.$

Therefore EA is invariant.

The quotient X/EA has two objects, A and B, with quotient dynamics

$A \leftrightarrow B.$

Now define

EB =
{{1, 2}, {3, 4}}.

Each class is preserved by T:

$T(\{1, 2\}) = \{1, 2\},$

T({3, 4}) = {3, 4}.

Therefore EB is also invariant.

Its quotient has two fixed objects:

C → C,

D → D.

The policies EA and EB have equal cardinal complexity.

They are incomparable under refinement.

EA groups corresponding positions in the two cycles.

EB groups states by cycle membership.

Transformation invariance alone does not select between them.

The task determines which quotient is useful.

Suppose the observable hphase records position within a cycle:

hphase(1) = 0,

hphase(3) = 0,

hphase(2) = 1,

hphase(4) = 1.

Then EA has zero observable distortion because hphase is constant on its classes.

EB does not.

Each EB-class contains one state of phase 0 and one of phase 1.

Thus EA is the natural phase quotient.

Now suppose the observable hcycle records which cycle contains the state:

hcycle(1) = 0,

hcycle(2) = 0,

hcycle(3) = 1,

hcycle(4) = 1.

Then EB has zero distortion and EA does not.

The same dynamics therefore support two equally small invariant object systems.

One preserves temporal phase.

The other preserves component identity.

There is no contradiction.

They are optimized for different policies π.

A third quotient is the universal relation

∇X = {{1, 2, 3, 4}}.

It is also invariant.

Its quotient has one fixed object.

It discards both phase and cycle identity.

The equality relation is invariant as well and retains all four states.

The invariant-policy lattice therefore contains multiple resolutions:

four detailed states,

two phase objects,

two cycle objects,

and one universal object.

Task distortion and complexity select among them.

This example also demonstrates that increasing pressure for simplicity need not generate a single nested hierarchy.

EA and EB have the same complexity but organize the state space differently.

A change in task can move the optimum between incomparable policies without changing the number of objects.

This is an ontological reorganization rather than a simple fusion or split.

The same principle extends to stochastic dynamics.

Let P(y | x) be a transition probability.

An equivalence relation E is exactly lumpable when equivalent states induce identical probabilities over quotient classes.

For every x E y and every quotient class C,

Σz∈C P(z | x)

Σz∈C P(z | y).

When this condition holds, the quotient process is Markovian.

If the condition holds only approximately, the variation in these class-transition probabilities defines a stochastic invariance defect.

The variational problem can then choose between:

a fine exactly lumpable quotient,

a coarse approximately lumpable quotient,

or a richer quotient model carrying memory.

The preferred solution depends on the cost assigned to state complexity, transition complexity, and non-Markovian distortion.

11.3 Multisensor Object Formation

Consider three sensors observing overlapping portions of one environment.

Sensor 1 observes states p and q and distinguishes them:

U1 = {p, q},

E1 = {{p}, {q}}.

Sensor 2 observes p and r and declares them equivalent:

U2 = {p, r},

E2 = {{p, r}}.

Sensor 3 observes r and q and declares them equivalent:

U3 = {r, q},

E3 = {{r, q}}.

No sensor directly asserts both

p equivalent to q

and

p distinct from q.

The local declarations may therefore appear pairwise manageable.

Globalization changes the result.

The union of local identity declarations contains

p E2 r,

r E3 q.

Equivalence closure forces

p Eglobal q.

Restricting the global relation back to U1 gives

Eglobal|U1 = {{p, q}}.

This is strictly coarser than E1.

Therefore

$C\mathcal{U}(\mathcal{E}) \neq \mathcal{E}$.

Faithful descent fails.

The global identity path

p → r → q

returns to Sensor 1 and fuses two locally distinct objects.

The local-object graph makes the defect visible.

Sensor 1 contributes nodes

A = [p]E1,

B = [q]E1.

Sensor 2 contributes one node

C = [p]E2 = [r]E2.

Sensor 3 contributes one node

D = [r]E3 = [q]E3.

Shared-state edges connect:

A to C through p,

C to D through r,

D to B through q.

All four local object nodes lie in one connected component.

That component contains both A and B from Sensor 1.

The no-collision condition fails.

Several repairs are possible.

First, accept global fusion.

Replace E1 by

E1′ = {{p, q}}.

The local family becomes closure-coherent, but Sensor 1 loses its distinction.

This is appropriate only when the distinction was uncertain or unnecessary.

Second, reject the identity declaration p ~ r.

Then the path breaks.

Third, reject the declaration r ~ q.

Again, the path breaks.

Fourth, refine one of the local policies using new sensor evidence.

Fifth, add another observation capable of determining which bridge is false.

Suppose the two cross-sensor identity bridges carry confidence weights:

w(p ~ r) = 0.9,

w(r ~ q) = 0.6.

Suppose Sensor 1’s separation between p and q is hard.

The maximum-weight faithful solution retains p ~ r and rejects r ~ q.

The retained correspondence weight is 0.9.

The alternative retains r ~ q and rejects p ~ r, with weight 0.6.

Thus the variational repair selects the first option.

If the separation were soft, the objective could compare:

the cost of rejecting one identity bridge,

against

the cost of merging Sensor 1's two objects.

A high descent penalty preserves local separation.

A low descent penalty permits global fusion.

This model applies directly to multisensor tracking.

Two cameras may each maintain distinct tracks.

Cross-camera correspondence edges can connect the tracks through intermediate observations.

A cycle of associations can cause two tracks visible simultaneously in one camera to become one global identity.

The local simultaneity evidence is then a protected distinction.

The correct global identity system must preserve it.

The example also applies to record linkage, knowledge graphs, biological specimen tracking, and distributed scientific observation.

Any identity system formed by transitive closure can generate remote fusion.

Local pairwise accuracy is not sufficient.

The complete path structure must be audited.

11.4 Scientific Taxonomy

A scientific taxonomy is an identity policy imposed on a set of specimens, states, mechanisms, or phenomena.

Different taxonomies arise because different scientific tasks preserve different structures.

Let X be a finite set of biological specimens.

Suppose three maps are available:

$m: X \to M$,

assigning morphological type,

$g: X \to G$,

assigning genetic lineage,

and

$r: X \to R$,

assigning intervention or ecological response.

The morphology-preserving identity policy is

$E_m = \mathrm{Ker}(m)$.

Two specimens are equivalent when their measured morphology agrees.

The genetic policy is

$E_g = \mathrm{Ker}(g)$.

The response policy is

$E_r = \mathrm{Ker}(r)$.

These relations need not coincide.

Two specimens may be morphologically similar but genetically distant.

Two genetically close specimens may have different ecological responses.

Two mechanistically distinct organisms may respond similarly under one intervention.

A taxonomy preserving all three structures exactly must satisfy

$E \subseteq E_m \cap E_g \cap E_r$.

The coarsest exact policy is

$E_{exact} = E_m \cap E_g \cap E_r$.

This relation can be much finer than any single-feature classification.

It retains every distinction made by at least one required scientific map.

Suppose morphology is inexpensive to observe, while genetic and response data are costly.

A practical taxonomy may tolerate limited genetic or response distortion in exchange for fewer classes and easier field recognition.

Define

$D(E) =$
$\alpha D_m(E)$
$+$

βDg(E)
+
γDr(E),

where the weights represent scientific priorities.

The complexity term may include more than class count.

A taxonomy is operationally complex when class membership requires expensive tests.

Let

K(E) =
λ1 |X/E|
+
λ2 test-cost(E)
+
λ3 description-length(E).

A taxonomy with many classes defined by visible morphology may be easier to use than a smaller taxonomy requiring whole-genome analysis.

The variational principle can capture this distinction.

Taxonomic revision then appears as policy movement in Eq(X).

A class splits when new genetic or causal evidence reveals significant internal distortion.

Classes merge when distinctions prove irrelevant to the selected scientific purpose.

The task policy must be stated explicitly.

A conservation taxonomy may preserve reproductive isolation and ecological vulnerability.

A pharmacological taxonomy may preserve drug response.

An evolutionary taxonomy may preserve ancestry.

A field-identification taxonomy may prioritize observable morphology.

These policies can generate different optimal object systems from the same specimens.

Scientific disagreement may therefore arise from different preservation objectives rather than from different raw evidence.

The variational framework makes those objectives visible.

A robust scientific taxon would be one that remains unified under several independent maps:

morphology,

genetics,

development,

ecology,

intervention response,

and transformation through life stages.

Such a class occupies a stable region across several policies π.

It has stronger objecthood than a category supported by one superficial feature alone.

Taxonomy also illustrates explanatory complexity.

A classification may reduce the number of groups but require a separate exception list for many specimens.

Its apparent cardinal compression can increase total descriptive complexity.

A better taxonomy may use more classes but explain each through a concise mechanism.

Thus the best scientific object system minimizes the combined cost of:

class count,

membership rule,

predictive error,

mechanistic inconsistency,

and exceptions.

11.5 Image Recognition and Transformation Invariance

Let X be a set of images.

Each image contains variations in:

position,

orientation,

scale,

illumination,

background,

occlusion,

and sensor noise.

A recognition system should often treat many of these variations as preserving one object.

Let G be a family of nuisance transformations acting on images.

For an exact group action, define the orbit of x by

$Orb_G(x) = \{g(x) : g \in G\}$.

The orbit relation is

$x \; E_G \; y$

when y belongs to $Orb_G(x)$.

If G is a group, E_G is an equivalence relation.

Its classes are transformation orbits.

The quotient

X/E_G

represents images up to the transformations in G.

For example, if G contains rotations and translations, one quotient object contains all rotated and translated presentations of one image pattern.

The quotient is exactly invariant under G.

Applying another transformation does not leave the orbit.

This is a mathematically clean form of recognition invariance.

Real image recognition is more difficult.

Not every nuisance transformation is invertible.

Occlusion can destroy information.

Illumination changes may saturate a sensor.

Background alteration may introduce new objects.

Learned deformations may not form a group.

The generated relation may connect images through chains that should not all be treated as one semantic object.

Thus transformation closure alone can overmerge.

A semantic observable

$h: X \to Y$

may assign the depicted object identity.

The admissible relation must preserve h:

x E y implies h(x) = h(y).

It must also be stable under nuisance transformations:

x E y implies g(x) E g(y)

for every $g \in G$.

The optimal recognition policy can therefore be written

$E\bigstar \in \arg\min \{$
Dsemantic(E)
+
λK(E)
+
vSinv(E)
:
E preserves protected semantic distinctions
}.

The distortion term penalizes classes containing different depicted objects.

The complexity term rewards invariance and representation compression.

The invariance term penalizes failure to recognize transformed versions as one quotient object.

A simple finite construction illustrates the tradeoff.

Suppose X contains twelve images:

three semantic objects,

each photographed in four orientations.

The equality policy has twelve classes.

It has zero semantic distortion but no recognition compression.

The orientation-orbit policy has three classes.

It has zero semantic distortion if every orientation preserves object identity.

The universal policy has one class.

It has minimal complexity but merges the three semantic objects and therefore has high distortion.

The three-class orbit quotient is the natural optimum across a broad range of λ.

Now suppose one object changes meaning under orientation, as can occur with certain symbols.

Then full rotational invariance becomes invalid for that object.

The optimal policy may use object-dependent transformation families.

This shows that admissible transformations cannot always be imposed globally.

A policy-indexed transformation family may be required.

Image recognition also demonstrates the difference between similarity and identity.

Two photographs of the same object under different lighting may have large pixel distance.

Two different objects may have small pixel distance.

The useful equivalence relation is determined by stable task behavior, not raw geometry.

A learned embedding can help estimate that relation, but the induced classes must still be tested for:

semantic distortion,

transformation stability,

and protected distinctions.

An adversarial perturbation is particularly informative.

If two images lie in one quotient class but a small allowed perturbation sends them to different semantic predictions, the quotient is unstable.

The invariance defect provides a certificate that the current image object is not robust under the declared transformation regime.

11.6 Distributed Identity and Descent Penalties

Consider four local domains with quotient objects:

Observer 1: A1, A2

Observer 2: B1, B2

Observer 3: C1

Observer 4: D1, D2.

Suppose candidate cross-domain identity edges have weights:

A1—B1 with weight 0.9,

B1—C1 with weight 0.8,

C1—D1 with weight 0.9,

A2—B2 with weight 0.7,

B2—C1 with weight 0.6,

C1—D2 with weight 0.7.

If all edges are accepted, one connected component contains

A1, A2, B1, B2, C1, D1, D2.

This component contains two objects from Observer 1, two from Observer 2, and two from Observer 4.

Faithful descent fails in three domains.

Define the retained-edge benefit

B(S) = sum of weights of accepted edges in S.

Define a collision penalty

$L_{coll}(S) = \sum_i \sum_{components\ Q} \max(0, m_i(Q) - 1)$,

where $m_i(Q)$ counts objects from observer i in component Q.

Let the optimization maximize

$V(S) = B(S) - \mu L_{coll}(S)$.

When μ is small, retaining all correspondence evidence may be optimal despite local fusion.

When μ is large, the selected graph must separate same-observer objects.

Suppose descent is hard.

Then no component may contain both A1 and A2, both B1 and B2, or both D1 and D2.

The optimization must choose which bridges through C1 to retain.

One feasible solution retains

A1—B1,

B1—C1,

C1—D1.

Its total weight is

0.9 + 0.8 + 0.9 = 2.6.

The second branch can retain

A2—B2

and

C1—D2,

but it cannot also retain B2—C1 because C1 is already connected through B1 and would fuse B1 with B2.

One possible feasible retained set is therefore

A1—B1,

B1—C1,

C1—D1,

A2—B2,

with total weight

3.3.

Another feasible solution could choose the second branch through C1:

A2—B2,

B2—C1,

C1—D2,

plus A1—B1,

with total weight

2.9.

The first feasible solution has greater retained weight.

It identifies A1, B1, C1, and D1 as traces of one global object.

It identifies A2 and B2 as another global object.

D2 remains separate.

The rejected edges are not simply low-confidence errors considered independently.

They are rejected because their inclusion would create prohibited global collisions after closure.

This is the defining nonlocal feature of descent optimization.

The marginal value of B2—C1 depends on which other edges have already been accepted.

If B1—C1 were absent, B2—C1 might be admissible.

Identity evidence cannot therefore be evaluated pair by pair in isolation.

The connected-component consequences must be recomputed.

A penalized version can permit one collision when its correspondence benefit is sufficiently high.

For example, adding B2—C1 to the first solution adds weight 0.6 but creates at least one collision in Observer 2 and one in Observer 1 through the larger component.

If μ exceeds the benefit per induced collision, the edge is rejected.

If μ is sufficiently small, the system accepts fusion.

This model provides a direct algorithmic interpretation of descent penalties.

The nodes are local quotient objects.

The decision variables are cross-domain correspondence edges.

Equivalence closure is graph connectivity.

Faithful descent is a color constraint: no component may contain two nodes of the same observer color.

The optimization selects the highest-value colored-component structure subject to or penalized by collisions.

This construction is applicable to:

multicamera tracking,

database entity resolution,

distributed biological specimen records,

multilingual knowledge graphs,

and federated scientific observations.

11.7 Emergent Macrostates

Consider a system of n binary components.

The microscopic state space is

$X = \{0, 1\}^n$.

It contains

$|X| = 2^n$

microstates.

Define the macro-observable

m(x) = number of components equal to 1.

The possible values are

0, 1, 2, …, n.

Define the equivalence relation

x Em y

when

m(x) = m(y).

The quotient has

$|X/Em| = n + 1$

macrostates.

The compression is substantial.

The detailed state space grows exponentially with n.

The macrostate space grows linearly.

For every observable h depending only on m,

$h(x) = \hat{h}(m(x))$,

the quotient has zero distortion.

The map h factors through X/Em.

Thus Em is the coarsest exact identity policy preserving the count m.

The equivalence classes have sizes

$|[x]_{E_m}|$ = number of binary strings with m(x) ones.

States with the same count may differ in the locations of their active components.

Those positional differences are erased.

Whether this is lawful depends on the dynamics.

Suppose the transformation family is invariant under permutations of component positions and the transition probabilities depend only on the current count m.

Then Em can support autonomous quotient dynamics.

The probability of moving from count k to count l depends only on k, not on the particular microstate within the class.

The macrostate process is exactly lumpable.

The quotient objects are dynamically lawful.

Now suppose interactions depend on spatial adjacency.

Two microstates with the same number of active components may have different future behavior because one contains a large cluster and another contains isolated active sites.

Then Em has dynamical distortion.

The macrostate count alone is insufficient.

A refined policy might preserve both:

the total count,

and the number of adjacent active pairs.

The quotient becomes more complex but more dynamically autonomous.

This is the variational construction of a better macrostate.

The theory does not assume that one familiar aggregate is correct.

It asks which equivalence relation produces the best balance between macrostate simplicity and dynamical fidelity.

A coarser policy can group counts into ranges.

For example,

Elow-high =
{
states with m < n/2,
states with m = n/2,
states with m > n/2
}.

This quotient has at most three objects.

It may preserve only the sign of deviation from half occupancy.

Its complexity is far lower than n + 1, but its distortion is greater for tasks requiring precise m.

A still coarser policy may retain only whether m exceeds a threshold.

The phase-transition path can therefore move through:

2^n microstates,

n + 1 exact count macrostates,

three sign macrostates,

two threshold macrostates,

and one universal state.

Each level is appropriate for a different distortion tolerance and task.

This example gives a precise interpretation of emergence.

The macrostate k is not an additional microscopic configuration.

It is the equivalence class of all configurations with k active components.

It becomes an effective object when:

the compression is valuable,

the relevant observables are nearly constant on the class,

and the dynamics approximately close on the quotient.

The same construction applies to:

particle-number macrostates,

energy shells,

magnetization levels,

population counts,

network degree summaries,

and aggregate economic states.

The correct macro-object is determined by the transformations and predictions it must support.

A coarse variable that does not induce autonomous behavior may still be descriptively useful, but it is not a complete dynamical object.

The invariance defect quantifies the missing closure.

11.8 Computational Search for Optimal Quotients

A finite implementation can search directly over identity policies.

The procedure begins with five inputs:

a finite state space X,

task data defining $D\pi$,

a complexity functional $K\pi$,

a transformation family G,

and structural constraints such as protected pairs or local domains.

For small X, exhaustive search is possible.

Generate every partition $E \in Eq(X)$.

For each E:

construct its quotient classes,

test hard constraints,

calculate distortion,

calculate complexity,

test transformation invariance,

calculate descent closure if distributed,

and evaluate the total objective.

The global minimum and all tied minimizers can then be reported exactly.

The number of partitions grows rapidly, so exhaustive enumeration is limited to small systems. Its value is nevertheless substantial.

It can validate approximate algorithms.

It can reveal all phase transitions.

It can identify nonunique optima.

It can test conjectures about monotonicity or lattice structure.

It can provide exact benchmark instances.

For larger systems, a merge-split search is more practical.

A basic agglomerative procedure is:

1. Begin with equality $E = \Delta X$.
2. Construct the current quotient classes.
3. Generate admissible candidate class mergers.
4. For each merger, compute the change in total objective.
5. Reject mergers violating hard constraints.
6. Accept the merger with the greatest objective improvement.
7. Update class statistics, transformation images, and distributed components.
8. Repeat until no merger improves the objective.

A divisive procedure begins with a coarse policy and uses defect-driven splits.

A bidirectional procedure alternates merges and splits.

Several computational shortcuts are available.

Classwise sufficient statistics can permit fast distortion updates.

If distortion depends on class label counts, merging two classes requires only their count vectors.

If dynamics are finite, store the quotient image distribution of every class under every generator.

A merge or split then updates only affected entries.

For distributed identity, union-find data structures maintain connected components efficiently.

However, standard union-find supports merging better than splitting. Dynamic graph methods or rollback structures may be needed when identity edges are removed.

Protected distinctions can be checked by maintaining observer-color sets within each component.

A proposed merge is infeasible if the two components contain prohibited duplicate colors or protected pairs.

The output of a computational search should include:

the optimal policy $E^\star$,

the quotient classes,

$D\pi(E^\star)$,

$K\pi(E^\star)$,

the total objective value,

the active constraints,

the induced quotient dynamics,

all detected descent and invariance defects,

alternative minimizers,

the objective gap to the next-best policy,

and critical λ values.

The quotient alone is not enough.

The supporting certificate makes the result auditable.

A phase scan evaluates the optimization across a sequence of λ values.

Because the finite objective is piecewise affine, a more exact method computes pairwise exchange values among candidate frontier policies.

The resulting diagram shows:

where objects merge,

where classes split as λ decreases,

where incomparable policies exchange optimality,

and where degeneracy occurs.

A computational system should also test robustness to uncertainty.

Suppose distortion values are estimated from data.

Let

$\hat{D}\pi(E)$

be the estimated distortion and

$\delta(E)$

an uncertainty bound.

A policy should be considered securely optimal only when its objective advantage exceeds the possible estimation error.

If two policies differ by less than their uncertainty bounds, the correct output is ontological ambiguity, not an unjustified unique selection.

The system can then choose an active observation designed to distinguish them.

A full discovery loop is:

construct an initial candidate policy,

optimize over feasible merges and splits,

identify near-optimal alternatives,

locate the state pairs on which those alternatives disagree,

select a measurement or intervention targeting those pairs,

update the distortion and constraint estimates,

and reoptimize.

This transforms object formation into an experimental computational science.

The algorithm does not merely assign names to states.

It proposes substitution laws, tests their consequences, and revises them when they fail.

The eight constructions in this section establish the breadth of the framework.

Finite classification shows how task labels and protected distinctions interact.

Dynamical lumping shows that invariance permits several task-dependent quotients.

Multisensor formation shows how global closure can destroy local objecthood.

Scientific taxonomy shows that object classes depend on which biological or explanatory structures must be preserved.

Image recognition shows how nuisance transformations generate useful but potentially dangerous invariances.

Distributed graph optimization shows how identity evidence must be evaluated through its global closure consequences.

Macrostate construction shows how emergence can be represented as lawful compression of microscopic states.

Computational search shows how the theory can be implemented and audited on finite systems.

The next section combines these elements into a general variational objecthood theorem and a broader research program. It states the unified constrained optimization result, clarifies its assumptions and limitations, and identifies the mathematical and computational developments required for a full theory of discovered objecthood.

12. Variational Objecthood Theorem and Research Program

The preceding sections construct a general mathematics of object selection.

A state space supplies possible distinctions.

An equivalence relation determines which distinctions are erased.

The quotient supplies the resulting object system.

Task distortion measures the consequences of erasure.

Complexity measures the burden of retaining distinctions.

Transformation invariance determines whether quotient objects support representative-independent operations.

Closure coherence determines whether distributed local identities survive globalization.

The variational principle selects among the policies satisfying these requirements.

The theory can now be summarized in one statement:

Effective objects are the equivalence classes of an identity policy that optimally balances task-relevant preservation against effective complexity while satisfying the required laws of transformation and local-to-global coherence.

This statement contains both a structural and a variational component.

The structural component determines which identity policies are lawful.

The variational component determines which lawful policy is preferred.

Neither component is sufficient alone.

Without structural constraints, optimization may select a small but dynamically ambiguous or globally incoherent quotient.

Without optimization, the theory may certify many lawful object systems without explaining why one resolution should be chosen.

The complete framework therefore has four layers:

candidate identity,

structural admissibility,

task evaluation,

and variational selection.

This section states the combined theorem for finite systems, clarifies its global and distributed forms, develops the relation between fixed points and invariant policies, summarizes compression-distortion optimality, and identifies the limits of scalar objecthood functionals. It

concludes with a research program for extending variational objecthood to infinite state spaces, categories, sheaves, homotopy-coherent structures, learned representations, and autonomous scientific discovery.

12.1 The Combined Variational Objecthood Theorem

Let X be a finite nonempty state space.

Let

Eq(X)

be the complete lattice of equivalence relations on X, ordered by refinement:

$E \preceq F$

when

$E \subseteq F$.

Thus E is finer than F.

Let G be a family of admissible transformations on X.

Define

Inv(G) =
{E ∈ Eq(X) :
x E y implies g(x) E g(y)
for every g ∈ G}.

Let H be a family of observables or task maps required to be preserved exactly.

Define

Pres(H) =
{E ∈ Eq(X) :
x E y implies h(x) = h(y)
for every h ∈ H}.

Let P be a set of protected unequal pairs.

Define

Sep(P) =
{E ∈ Eq(X) :

x not E y
for every (x, y) ∈ P}.

If the theory is distributed over a cover $\mathcal{U}$, let the policy variable be a local family

$\mathcal{E}$ = (Ei)i∈I

belonging to the product lattice

L$\mathcal{U}$ = ∏i∈I Eq(Ui).

Let C$\mathcal{U}$ be the globalization–restriction closure operator.

Define

Fix(C$\mathcal{U}$) =
{$\mathcal{E}$ ∈ L$\mathcal{U}$:
C$\mathcal{U}$($\mathcal{E}$) = $\mathcal{E}$}.

For a purely global problem, the closure condition may be omitted or replaced by a suitable global coherence constraint.

Let Aπ be the nonempty feasible family obtained by intersecting the required structural conditions.

In a global formulation, one may write

Aπ =
Inv(G)
∩
Pres(H)
∩
Sep(P)
∩
Aadditional.

In a distributed formulation,

Aπ,$\mathcal{U}$ =
Fix(C$\mathcal{U}$)
∩
Inv$\mathcal{U}$(G)
∩
Pres$\mathcal{U}$(H)
∩

$\mathrm{Sep}_{\mathcal{U}}(P)$
∩
Aadditional.

Let

D_π

be a nonnegative task-relevant distortion functional.

Let

K_π

be a nonnegative effective complexity functional.

For $\lambda \geq 0$, define the objecthood functional

$J_{\pi,\lambda}(E) =$
$D_\pi(E) + \lambda K_\pi(E)$

in the global case, or

$J_{\pi,\lambda}(\mathcal{E}) =$
$D_\pi(\mathcal{E}) + \lambda K_\pi(\mathcal{E})$

in the distributed case.

Variational Objecthood Theorem.

Under the assumptions above:

1. At least one minimizing identity policy exists.
2. Every minimizing policy produces a quotient object system satisfying all hard structural constraints.
3. The minimizer need not be unique.
4. The minimizer set contains at least one refinement-minimal and one refinement-maximal policy.
5. If a unique minimizer exists at a noncritical λ, it remains optimal under sufficiently small perturbations of λ and sufficiently small objective-estimation errors.
6. As λ increases, the complexity of every selected optimum weakly decreases and its distortion weakly increases.

7. The selected policies need not form a refinement chain as λ changes.
8. Changes in the minimizer family occur only at finitely many critical λ values.
9. At critical values, object fusion, object splitting under reverse parameter motion, or reorganization among incomparable policies may occur.
10. If the exact feasible family is empty, no identity policy satisfies all stated laws simultaneously.

Proof.

The candidate lattice or product lattice is finite because X and the local domains are finite.

Every feasible family is therefore finite.

A real-valued functional on a finite nonempty set attains its minimum.

Every minimizer lies in the feasible family and therefore satisfies the hard constraints by definition.

The minimizer set is a finite partially ordered set and consequently contains refinement-minimal and refinement-maximal elements.

The remaining claims follow from the affine dependence

$J_{\pi,\lambda}(E) = D_{\pi}(E) + \lambda K_{\pi}(E)$,

the finite lower-envelope argument, the objective-gap stability result, and the monotonicity inequalities established in Section 8.

The theorem is intentionally finite.

It does not claim that every infinite objecthood problem has a minimizer.

It does not claim uniqueness.

It does not claim that the selected quotient is independent of π.

It does not claim that every object system forms a nested hierarchy.

Its strength is more precise.

It shows that finite object discovery can be formulated as a complete mathematical extremization problem with exact solutions, explicit structural constraints, measurable tradeoffs, and auditable transition points.

A corresponding penalized theorem follows immediately.

Let

$$\Phi\pi(E) = D\pi(E) + \lambda K\pi(E) + \mu L\mathrm{des}(E) + \nu S\mathrm{inv}(E) + \rho L\mathrm{sep}(E) + \tau L\mathrm{revision}(E).$$

If every term is finite-valued on Eq(X), then Φπ has at least one minimizer.

Such a minimizer is an optimal compromise under the chosen penalties.

It is not necessarily an exact lawful object system.

The distinction between exact and penalized objecthood must remain explicit.

12.2 Constrained Quotient Optimization

The theorem is best understood as optimization over quotient-producing policies rather than direct optimization over unlabeled quotient sets.

For every $E \in \mathrm{Eq}(X)$, the quotient

$$X/E$$

is determined up to canonical isomorphism.

The policy E contains more information than the bare cardinality of the quotient.

It records which states become interchangeable.

Two equivalence relations can produce quotients with the same number of classes while representing entirely different ontologies.

Thus the optimization variable should normally be E rather than only X/E or $|X/E|$.

The master global problem is

$$E\star \in \arg\min \{ D\pi(E) + \lambda K\pi(E) : E \in A\pi \}.$$

The selected object space is then

$$O\star = X/E\star.$$

The quotient map is

$$q\star: X \to O\star$$

with

$$q\star(x) = [x]E\star.$$

Every task map preserved by $E\star$ factors through $q\star$.

Every admissible transformation preserving $E\star$ descends through $q\star$.

The quotient object space is therefore not merely a set of clusters.

It is the codomain of a universal compression map supporting the selected observables and operations.

The exact least-complexity problem is

$$E\star \in \arg\min \{ K\pi(E) : E \in A\pi \text{ and } D\pi(E) = 0 \}.$$

This asks for the simplest lawful quotient that loses none of the required task structure.

The bounded-distortion form is

$$E\star \in \arg\min \{ K\pi(E) : E \in A\pi \text{ and } D\pi(E) \leq \varepsilon \}.$$

The bounded-complexity form is

$$
\begin{aligned}
&E^\star \in \arg\min \{ \\
&D_\pi(E) : \\
&E \in A_\pi \\
&\text{and} \\
&K_\pi(E) \le B \\
&\}.
\end{aligned}
$$

The penalized form is

$$
\begin{aligned}
&E^\star \in \arg\min \{ \\
&D_\pi(E) \\
&+ \\
&\lambda K_\pi(E) \\
&+ \\
&\text{structural penalties} \\
&: \\
&E \in Eq(X) \\
&\}.
\end{aligned}
$$

These formulations produce related but not always identical minimizers.

The policy lattice is discrete.

A weighted sum may skip some Pareto-optimal policies.

A distortion bound may select a policy that is never optimal for any λ.

For this reason, a complete analysis should examine:

the weighted path,

the distortion-constrained path,

the complexity-constrained path,

and the full Pareto frontier.

Quotient optimization may also include the complexity of induced laws.

A more complete functional is

$$
\begin{aligned}
&J_\pi(E) = \\
&D_\pi(E) \\
&+
\end{aligned}
$$

λ1Kobjects(E)
+
λ2Kpolicy(E)
+
λ3Kdynamics(E)
+
λ4Kexplanation(E).

This prevents false compression.

A quotient with few objects may require a highly complicated equivalence test or non-Markovian effective dynamics.

Its object count is low, but its total theory may not be simple.

The correct quantity is effective complexity:

the total cost of defining the identity policy, naming the quotient objects, and representing their lawful behavior.

The optimization can also include a prior policy E0:

Jπ(E; E0) =
Dπ(E)
+
λKπ(E)
+
τdEq(E, E0).

The parameter τ measures ontology-revision cost.

This formulation is appropriate for adaptive systems, scientific taxonomies, databases, and AI memories in which conceptual change has operational consequences.

The constrained quotient problem therefore unifies several familiar tasks:

state aggregation,

model reduction,

sufficient representation,

entity resolution,

concept discovery,

taxonomy formation,

and scientific abstraction.

Its distinguishing feature is that the optimized variable is the law of identity itself.

12.3 Closure Fixed Points and Invariant Policies

Closure coherence and transformation invariance play different but complementary roles.

Closure coherence governs distributed composition.

Transformation invariance governs temporal, operational, or algebraic persistence.

A distributed policy family $\mathcal{E}$ is closure-coherent when

$C\mathcal{U}(\mathcal{E}) = \mathcal{E}$.

An identity policy E is transformation-invariant when

x E y implies g(x) E g(y)

for every admissible g.

The first condition says:

Global integration adds no locally forbidden identity.

The second says:

Admissible transformation does not separate states already declared identical.

Together they establish two directions of stability.

Closure coherence protects identity across space, context, and observer integration.

Transformation invariance protects identity across action, time, and operation.

A lawful distributed object should survive both.

Let

Acoh = Fix($C\mathcal{U}$),

and

Adyn = Inv(G).

The structurally admissible region is

Astruct = Acoh ∩ Adyn

with the appropriate product-lattice interpretation.

The intersection is closed under arbitrary meets when both families are meet-closed.

This means that common refinement is structurally safe.

If several policies are coherent and invariant, their intersection remains coherent under suitable distributed formulation and remains invariant.

Coarsening is more delicate.

The ambient join may require closure under both descent and transformation consequences.

One may define:

Cdes = descent closure,

Cinv = invariant equivalence closure.

A common closure can be constructed by iterating the two operators:

E0 = E,

E1 = Cdes(Cinv(E0)),

E2 = Cdes(Cinv(E1)),

and so forth.

On a finite lattice, an extensive monotone iteration stabilizes.

The limit is a common coarsening closed under both systems of consequences.

This construction gives the least common completion under the chosen operator order when the relevant closure-theoretic conditions hold.

However, closure completion can create excessive fusion.

Adding one identity to repair transformation stability may create a descent collision.

Repairing the descent collision by further coarsening may create new task distortion.

The closures therefore cannot be applied blindly.

Their consequences must be evaluated by the variational functional.

A dual approach repairs by refinement.

If E fails invariance, split unstable classes.

If $\mathcal{E}$ fails descent, remove identity bridges or refine local classes.

Refinement preserves more distinctions but increases complexity.

Coarsening preserves positive identity declarations but may erase protected distinctions.

The variational problem decides which direction of repair is less costly.

This yields a general repair principle.

Given a proposed policy E0, search for

$$E^{\bigstar} \in \arg\min \{ d_{Eq}(E, E_0) + \lambda K_{\pi}(E) + D_{\pi}(E) : E \in \mathrm{Astruct} \}.$$

The result is the nearest lawful ontology under the selected policy distance and task costs.

Fixed-point methods and variational methods therefore serve different functions.

Closure determines completion.

Optimization determines selection.

A fixed point is self-consistent.

An optimum is preferred.

A policy can be one without being the other.

A closure fixed point may be needlessly fine or destructively coarse.

A low-cost policy may be structurally invalid.

Variational objecthood requires both properties simultaneously.

12.4 Compression-Distortion Optimality

The central selection law is the competition between compression and distortion.

A quotient compresses by replacing many states with one effective object.

It distorts when the erased differences affect the selected task.

The two extreme policies reveal why both terms are necessary.

Equality gives maximal retained detail:

$E = \Delta X$.

It usually minimizes distortion.

It maximizes quotient resolution.

Universal equivalence gives maximal compression:

$E = \nabla X$.

It minimizes class-count complexity.

It usually maximizes task loss.

A useful object system lies between these extremes unless the task itself is constant or every distinction is essential.

For two policies E and F, suppose F is coarser:

$E \preceq F$.

Let

$\Delta D = D\pi(F) - D\pi(E)$,

and

$\Delta K = K\pi(E) - K\pi(F)$.

Under ordinary monotonicity,

$\Delta D \geq 0$,

$\Delta K \geq 0$.

The coarsening is favored when

$\Delta D < \lambda \Delta K$.

The distortion added by merging classes is smaller than the weighted complexity saved.

The critical exchange price is

$\lambda c = \Delta D / \Delta K$

when $\Delta K > 0$.

Below λc, the finer policy is preferred.

Above λc, the coarser policy is preferred.

This ratio assigns a precise value to a distinction.

A distinction survives while the task cost of erasing it exceeds the complexity cost of retaining it.

It disappears when the balance reverses.

The same equation applies to a proposed split.

A split is justified when its distortion reduction exceeds its weighted complexity increase.

Thus object fusion and object birth are governed by one local exchange law.

Global optimization remains necessary because local decisions interact.

Merging one pair can alter:

future merge costs,

transformation invariance,

descent connectivity,

class statistics,

and induced quotient dynamics.

The globally optimal ontology may therefore differ from the result of greedy local decisions.

The Pareto frontier provides the complete set of efficient policies.

A policy is meaningful for variational selection only when no alternative is both simpler and less distorted.

The frontier can contain:

nested partitions,

incomparable partitions,

cost-degenerate policies,

and unsupported nondominated points.

This means that conceptual resolution is not always one-dimensional.

There may be several distinct ways to simplify the same state space.

One policy preserves temporal behavior.

Another preserves causal response.

Another preserves morphology.

Another preserves explanation.

All may occupy different regions of a multidimensional frontier.

A scalar λ chooses one tradeoff direction.

A broader theory may require vector-valued complexity and distortion.

Nevertheless, the scalar compression-distortion law is the essential first principle.

It supplies the equation missing from a theory based only on quotient certification.

12.5 Existence and Multiplicity of Object Systems

Finite existence is guaranteed by the discreteness of the policy space.

Multiplicity is equally fundamental.

Let

$$M_{\pi,\lambda} = \arg\min \{J_{\pi,\lambda}(E) : E \in A_\pi\}.$$

The selected ontology is unique only when

$$|M_{\pi,\lambda}| = 1.$$

Several policies may tie because:

they have identical distortion and complexity,

they are related by symmetry,

they preserve different structures of equal task value,

the evidence is insufficient,

the objective is too coarse,

or λ lies at a critical exchange point.

Multiplicity should not automatically be removed.

It can represent genuine underdetermination.

A mathematical system should distinguish three cases.

In the first case, multiple policies are isomorphic under an admissible symmetry.

Their difference is representational.

In the second case, multiple policies are semantically distinct but task-equivalent.

The selected policy π does not care which one is used.

In the third case, multiple policies are distinct because current evidence cannot determine which is correct.

Further observation may break the tie.

These cases require different responses.

Symmetry-equivalent policies may be represented by one orbit of solutions.

Task-equivalent policies may be retained as interchangeable implementations.

Evidence-underdetermined policies should motivate active observation.

The meet of all minimizers,

Ecore = $\bigcap E \in M_{\pi,\lambda}$ E,

contains every identity accepted by all optimal policies.

It represents the robust identity core.

The join,

Eenvelope = EqCl($\bigcup E \in M_{\pi,\lambda}$ E),

contains every identity accepted by at least one optimum.

It represents the broad identity envelope.

Neither must itself be optimal.

The interval between them records policy uncertainty.

A pair of states equivalent in Ecore is robustly identical under the current optimization.

A pair separated in Eenvelope is robustly distinct.

Pairs lying between these extremes are ontologically unresolved.

This provides a useful three-way classification:

robust identity,

robust distinction,

and policy-dependent identity.

Tie-breaking can select one policy when necessary.

Possible secondary rules include:

finest optimal policy,

coarsest optimal policy,

shortest policy description,

simplest quotient dynamics,

closest policy to a prior ontology,

or most robust policy under perturbation.

Every such rule introduces additional assumptions.

It should be stated explicitly.

An algorithm returning one minimizer from a degenerate set does not establish unique objecthood.

The theory therefore favors reporting:

the minimizer set,

the objective gap,

the refinement relations among minimizers,

and the identities on which they disagree.

Multiplicity is not a defect of variational objecthood.

It is information about the limits of the current task and evidence.

12.6 Implications for Mathematics, Science, and Artificial Intelligence

The framework has several broad implications.

The first is that objecthood can be treated as a mathematical output.

Classical quotient theory begins with an equivalence relation.

Variational objecthood adds a principle for selecting the relation.

The quotient is no longer merely available.

It is justified by an extremal property.

The second implication is that identity is policy-indexed but not arbitrary.

Different tasks can produce different optimal equivalence relations.

The dependence is explicit.

One can identify which observable, intervention, transformation, or complexity price caused the ontology to change.

Policy dependence therefore does not imply unrestricted subjectivity.

It implies conditional mathematical objectivity.

Given X, π, G, the constraints, Dπ, Kπ, and λ, the minimizer set is determined.

The third implication is that abstraction can be quantified.

A good abstraction is not merely a simpler description.

It is a quotient whose lost distinctions have low task cost and whose retained structure supports lawful operations.

This provides a common language for:

coarse-graining,

state minimization,

scientific idealization,

concept formation,

and model reduction.

The fourth implication is that emergence can be formalized as optimized quotient change.

A macro-object appears when a coarser policy becomes preferable while retaining sufficient dynamical and explanatory autonomy.

An object subtype appears when refinement becomes necessary.

Emergence is therefore not only the presence of a pattern.

It is the variational selection of a new identity class.

The fifth implication is that distributed identity requires descent.

Entity resolution cannot be reduced to independent pairwise matches.

Transitive closure can create remote false fusion.

A global object must preserve locally authoritative distinctions or explicitly account for their revision.

The sixth implication is that object stability is operational.

An equivalence class is not a lawful object for a transformation regime unless the transformations descend to the quotient.

Static clustering does not guarantee dynamical objecthood.

The seventh implication is that scientific taxonomies can be compared through explicit preservation objectives.

Two taxonomies may disagree because they optimize different maps:

ancestry,

morphology,

mechanism,

prediction,

or intervention.

The framework makes the source of disagreement mathematically inspectable.

The eighth implication concerns artificial intelligence.

Most AI systems learn functions within a supplied object space.

A stronger system would learn the identity policy defining that object space.

It would monitor distortion inside its concepts.

It would split concepts whose members behave differently.

It would merge concepts whose distinction no longer matters.

It would preserve hard safety distinctions.

It would maintain task-indexed ontologies.

It would request observations capable of resolving uncertainty among competing policies.

Such a system would not merely learn representations.

It would revise its units of representation.

The ninth implication concerns explanation.

An object class is explanatorily strong when one concise model applies across its members.

The best ontology may therefore minimize the combined complexity of objects and laws.

A quotient with fewer objects but highly irregular dynamics may be scientifically inferior to a larger quotient governed by a simple theory.

The tenth implication is methodological.

Claims about objecthood should include an admissibility certificate.

A rigorous report should state:

the state space,

the selected equivalence relation,

the quotient classes,

the task policy,

the distortion measure,

the complexity measure,

the transformation family,

the descent conditions,

the active constraints,

the objective value,

and the alternative minimizers.

This converts conceptual classification into an auditable mathematical procedure.

12.7 Limits of Scalar Objecthood Functionals

The scalar objective

$D\pi(E) + \lambda K\pi(E)$

is powerful because it is compact, interpretable, and mathematically tractable.

It is not a complete theory of every form of objecthood.

Several limits must be acknowledged.

First, one scalar distortion value can conceal qualitatively different failures.

A policy may have modest average error while producing catastrophic error on one protected case.

Two policies can have equal D_π while one fails causally and the other fails predictively.

Separate distortion coordinates may be required.

Second, one scalar complexity value can conceal where complexity resides.

A quotient may have few objects but a complicated membership rule.

Another may have many objects governed by a simple algebra.

The total complexity should be decomposed.

Third, weighted sums assume compensability.

They assume enough compression can compensate for a given amount of distortion.

Some distinctions are nonnegotiable.

Logical contradiction, safety-critical identity, or certified disequality should often be represented as hard constraints rather than finite penalties.

Fourth, numerical weights can create an illusion of objectivity.

The values λ, μ, ν, and other coefficients encode priorities.

They must be justified by the task, resource regime, or decision context.

The optimization is only as meaningful as its policy specification.

Fifth, finite empirical distortion estimates may be wrong.

A class can appear homogeneous because the relevant transformation or intervention has not yet been observed.

Objecthood claims should include uncertainty and robustness analysis.

Sixth, equivalence relations impose exact transitivity.

Some real conceptual structures are graded, contextual, temporal, or nontransitive.

Similarity may not form an equivalence relation.

Approximate identity may require enriched structures:

fuzzy relations,

probabilistic partitions,

groupoids,

categories,

type-indexed equivalences,

or homotopy-coherent paths.

Seventh, a single global quotient may be inappropriate.

Different tasks may require different object systems.

The correct theory may be a family of quotients indexed by π rather than one universal partition.

Eighth, state spaces are often not fixed.

New observations can add states.

New features can refine what a state means.

A learned representation can change X itself.

The optimization may therefore need to vary both the state representation and the equivalence relation.

Ninth, quotient complexity can depend on representation language.

No fully universal description-length measure exists without assumptions about admissible encodings and primitives.

Representation invariance must be specified rather than presumed.

Tenth, optimizing over observed states may not generalize to unseen states.

A learned policy requires a rule extending equivalence beyond the finite sample.

The complexity and distortion of that rule must be evaluated out of sample.

Eleventh, the policy may affect the data used to evaluate it.

Once a system merges states, it may aggregate future observations and lose evidence needed for later refinement.

Ontology learning is therefore path-dependent and potentially self-reinforcing.

Twelfth, one optimum may not capture explanatory pluralism.

Several incompatible but useful object systems may coexist.

Forcing them into one scalar ranking can erase legitimate scientific alternatives.

These limits do not invalidate the variational principle.

They define its domain.

The scalar finite theory is the foundational case.

More expressive mathematics should extend it while preserving its central insight:

Objecthood is selected by lawful compression under explicit preservation conditions.

12.8 Next Steps in the Research Program

The first research direction is infinite-state existence theory.

For infinite X, Eq(X) is generally enormous and lacks the simple finite compactness used above.

One must identify topologies or order structures on admissible equivalence relations.

Existence may require:

compactness of the policy space,

lower semicontinuity of distortion,

coercivity of complexity,

closure of the feasible region,

or restricted parameterized policy families.

The goal is an infinite variational objecthood theorem analogous to the direct method of the calculus of variations.

The second direction is continuous and measurable quotient theory.

Physical and statistical state spaces often carry topology, geometry, measure, or smooth structure.

Not every set-theoretic equivalence relation produces a useful quotient.

One may require:

closed equivalence relations,

measurable equivalence relations,

smooth group actions,

or quotients satisfying separation properties.

Complexity may involve dimension, entropy, volume, or geometric regularity.

The third direction is categorical objecthood.

Equivalence relations on sets are only the simplest instance.

In categories, identity policies may be represented by congruences, reflective localizations, factorization systems, or classes of morphisms inverted by localization.

The variational variable may become a localization policy rather than a set partition.

The selected objects would be objects of an optimized quotient or localized category.

The fourth direction is homotopy-coherent identity.

Exact equivalence relations discard the structure of witnesses.

In many mathematical and physical settings, paths between states matter.

Different equivalence witnesses can compose in nontrivial ways.

Homotopy type theory and higher category theory provide richer identity structures.

Variational objecthood should eventually optimize not only which points are equivalent, but also which paths, higher paths, and coherence structures are retained.

The fifth direction is sheaf and stack descent.

The closure operator $C\mathcal{U}$ gives a set-level model of distributed gluing.

More structured objects require descent of:

functions,

groups,

categories,

probability models,

dynamical systems,

or explanatory mechanisms.

A variational stack theory would select globally coherent objects while minimizing local revision, structural complexity, and descent defect.

The sixth direction is vector-valued objecthood.

Instead of one distortion and one complexity, define vectors such as

D(E) =
(Dprediction,
Dcausal,
Ddynamical,
Dexplanatory,
Dsafety),

and

K(E) =
(Kobjects,
Kpolicy,
Kdynamics,
Kcomputation,
Kdescription).

Selection then occurs on a multidimensional Pareto frontier.

This avoids forcing every value into one scalar exchange rate.

The seventh direction is algorithmic scalability.

Exact partition search is practical only for small systems.

Larger problems require:

constraint propagation,

lattice pruning,

graph-based methods,

integer programming,

probabilistic partition models,

differentiable relaxations,

and hierarchical merge-split search.

Benchmark problems should include known optimal quotients so that approximate algorithms can be audited.

The eighth direction is uncertainty-aware objecthood.

Distortion, transformation laws, and identity evidence are usually estimated.

A robust theory should assign uncertainty to:

policy costs,

constraint validity,

and class membership.

The output may be a distribution over equivalence relations rather than one relation.

Active observation should be chosen to reduce uncertainty in the minimizer set.

The ninth direction is ontology dynamics.

Policies should be allowed to evolve:

$E_t \to E_{t+1}$.

The theory should study:

revision cost,

hysteresis,

memory loss,

irreversible coarsening,

and recovery of previously erased distinctions.

A dynamical equation on the policy lattice may complement the static variational principle.

The tenth direction is autonomous scientific AI.

A system implementing variational objecthood would:

generate candidate equivalence policies,

test their predictive and causal adequacy,

verify transformation stability,

check distributed descent,

estimate complexity,

retain near-optimal alternatives,

and design experiments that distinguish them.

Such a system would perform concept discovery rather than only pattern fitting.

The eleventh direction is robustness across policies.

An object may be task-specific or broadly stable.

Define a family Π of admissibility policies.

Study the region of Π under which a given identity remains optimal.

Objects stable across large regions may represent deeper structure.

This provides a mathematical route from policy-relative objecthood to robust objecthood.

The twelfth direction is empirical validation.

The theory should be tested on domains where quotient structure is independently understood:

finite automata,

Markov-state aggregation,

symmetry reduction,

multisensor tracking,

scientific taxonomies,

image invariance,

and physical macrostate discovery.

The aim is not merely to show that an objective can reproduce known partitions.

The stronger test is whether it predicts:

when known objects should merge,

when they should split,

which observations resolve ambiguity,

and which quotient laws remain autonomous.

The complete research program can be summarized by one equation:

$E^{\star} \in \arg\min \{ D_{\pi}(E) + \lambda K_{\pi}(E) : E \text{ satisfies the required laws of coherence, transformation, and preservation} \}.$

In the distributed case:

$\mathcal{E}^{\star} \in \arg\min \{ D_{\pi}(\mathcal{E}) + \lambda K_{\pi}(\mathcal{E}) : C\mathcal{U}(\mathcal{E}) = \mathcal{E} \text{ and } \mathcal{E} \text{ satisfies all transformation and preservation constraints} \}.$

These expressions are not proposed as universal numerical formulas independent of context.

They are schemas.

Each scientific domain must define:

what the states are,

which transformations matter,

which observations must be preserved,

which distinctions are protected,

how distortion is measured,

and what counts as complexity.

Once those choices are explicit, the theory determines the best lawful quotient or the complete set of equally optimal quotients.

The mathematical sequence is therefore:

X supplies states.

Eq(X) supplies possible identities.

Structural constraints define admissible identities.

D_{π} measures the cost of erased differences.

K_{π} measures the cost of retained distinctions.

Optimization selects $E^{\star}$.

The quotient $X/E^{\star}$ supplies the effective objects.

Parameter variation reveals their phase structure.

Active observation revises them.

This completes the transition from stable quotient theory to variational objecthood.

Objects are not merely assumed.

They are not merely named.

They are not merely certified after a policy is supplied.

They are selected as extremizers in a space of possible identity laws.

The resulting principle is compact:

Among all lawful ways of treating states as the same, choose the one that preserves what matters with the least effective complexity.

That principle provides the governing equation for the mathematics of discovered objecthood.

13. References

1. Birkhoff, G. 1935. On the structure of abstract algebras. Proceedings of the Cambridge Philosophical Society 31(4): 433–454. DOI: 10.1017/S0305004100013463.
2. Birkhoff, G. 1967. Lattice Theory. Third edition. American Mathematical Society Colloquium Publications, Volume 25. American Mathematical Society, Providence, Rhode Island.
3. Ore, O. 1942. Theory of equivalence relations. Duke Mathematical Journal 9(3): 573–627. DOI: 10.1215/S0012-7094-42-00942-6.
4. Rota, G.-C. 1964. On the foundations of combinatorial theory I: Theory of Möbius functions. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 2: 340–368. DOI: 10.1007/BF00531932.
5. Tarski, A. 1955. A lattice-theoretical fixpoint theorem and its applications. Pacific Journal of Mathematics 5(2): 285–309. DOI: 10.2140/pjm.1955.5.285.
6. Davey, B. A., and Priestley, H. A. 2002. Introduction to Lattices and Order. Second edition. Cambridge University Press, Cambridge. DOI: 10.1017/CBO9780511809088.
7. Burris, S., and Sankappanavar, H. P. 1981. A Course in Universal Algebra. Graduate Texts in Mathematics, Volume 78. Springer-Verlag, New York. DOI: 10.1007/978-1-4613-8130-3.
8. McKenzie, R. N., McNulty, G. F., and Taylor, W. F. 1987. Algebras, Lattices, Varieties. Volume I. Wadsworth & Brooks/Cole, Monterey, California.
9. Ganter, B., and Wille, R. 1999. Formal Concept Analysis: Mathematical Foundations. Springer-Verlag, Berlin. DOI: 10.1007/978-3-642-59830-2.
10. Mac Lane, S. 1998. Categories for the Working Mathematician. Second edition. Graduate Texts in Mathematics, Volume 5. Springer-Verlag, New York. DOI: 10.1007/978-1-4757-4721-8.
11. Mac Lane, S., and Moerdijk, I. 1992. Sheaves in Geometry and Logic: A First Introduction to Topos Theory. Springer-Verlag, New York. DOI: 10.1007/978-1-4612-0927-0.
12. Kashiwara, M., and Schapira, P. 2006. Categories and Sheaves. Grundlehren der Mathematischen Wissenschaften, Volume 332. Springer-Verlag, Berlin. DOI: 10.1007/3-540-27950-4.
13. Riehl, E. 2016. Category Theory in Context. Dover Publications, Mineola, New York.

14. Spivak, D. I. 2014. Category Theory for the Sciences. MIT Press, Cambridge, Massachusetts.

15. Curry, J. M. 2014. Sheaves, Cosheaves and Applications. Doctoral dissertation, University of Pennsylvania, Philadelphia.

16. Robinson, M. 2017. Sheaves are the canonical data structure for sensor integration. Information Fusion 36: 208–224. DOI: 10.1016/j.inffus.2016.12.002.

17. Ghrist, R. W. 2014. Elementary Applied Topology. CreateSpace Independent Publishing Platform.

18. The Univalent Foundations Program. 2013. Homotopy Type Theory: Univalent Foundations of Mathematics. Institute for Advanced Study, Princeton, New Jersey.

19. Shannon, C. E. 1948. A mathematical theory of communication. Bell System Technical Journal 27: 379–423 and 623–656.

20. Shannon, C. E. 1959. Coding theorems for a discrete source with a fidelity criterion. IRE National Convention Record 7, Part 4: 142–163.

21. Berger, T. 1971. Rate Distortion Theory: A Mathematical Basis for Data Compression. Prentice-Hall, Englewood Cliffs, New Jersey.

22. Cover, T. M., and Thomas, J. A. 2006. Elements of Information Theory. Second edition. Wiley-Interscience, Hoboken, New Jersey.

23. Rissanen, J. 1978. Modeling by shortest data description. Automatica 14(5): 465–471. DOI: 10.1016/0005-1098(78)90005-5.

24. Grünwald, P. D. 2007. The Minimum Description Length Principle. MIT Press, Cambridge, Massachusetts.

25. Tishby, N., Pereira, F. C., and Bialek, W. 1999. The information bottleneck method. Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing: 368–377.

26. Fisher, R. A. 1922. On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London, Series A 222: 309–368. DOI: 10.1098/rsta.1922.0009.

27. Casella, G., and Berger, R. L. 2002. Statistical Inference. Second edition. Duxbury, Pacific Grove, California.

28. Kemeny, J. G., and Snell, J. L. 1960. Finite Markov Chains. D. Van Nostrand, Princeton, New Jersey.

29. Buchholz, P. 1994. Exact and ordinary lumpability in finite Markov chains. Journal of Applied Probability 31(1): 59–75. DOI: 10.2307/3215235.

30. Larsen, K. G., and Skou, A. 1991. Bisimulation through probabilistic testing. Information and Computation 94(1): 1–28. DOI: 10.1016/0890-5401(91)90030-6.

31. Paige, R., and Tarjan, R. E. 1987. Three partition refinement algorithms. SIAM Journal on Computing 16(6): 973–989. DOI: 10.1137/0216062.

32. Milner, R. 1989. Communication and Concurrency. Prentice Hall, New York.

33. Li, L., Walsh, T. J., and Littman, M. L. 2006. Towards a unified theory of state abstraction for Markov decision processes. Proceedings of the Ninth International Symposium on Artificial Intelligence and Mathematics.

34. Jain, A. K., Murty, M. N., and Flynn, P. J. 1999. Data clustering: A review. ACM Computing Surveys 31(3): 264–323. DOI: 10.1145/331499.331504.

35. Wagstaff, K., Cardie, C., Rogers, S., and Schrödl, S. 2001. Constrained k-means clustering with background knowledge. Proceedings of the Eighteenth International Conference on Machine Learning: 577–584.

36. Bansal, N., Blum, A., and Chawla, S. 2004. Correlation clustering. Machine Learning 56(1–3): 89–113. DOI: 10.1023/B.0000033116.57574.95.

37. Meilă, M. 2007. Comparing clusterings—an information based distance. Journal of Multivariate Analysis 98(5): 873–895. DOI: 10.1016/j.jmva.2006.11.013.

38. Rand, W. M. 1971. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66(336): 846–850. DOI: 10.1080/01621459.1971.10482356.

39. Banerjee, A., Merugu, S., Dhillon, I. S., and Ghosh, J. 2005. Clustering with Bregman divergences. Journal of Machine Learning Research 6: 1705–1749.

40. Pearl, J. 2009. Causality: Models, Reasoning, and Inference. Second edition. Cambridge University Press, Cambridge. DOI: 10.1017/CBO9780511803161.

41. Spirtes, P., Glymour, C., and Scheines, R. 2000. Causation, Prediction, and Search. Second edition. MIT Press, Cambridge, Massachusetts.

42. Peters, J., Janzing, D., and Schölkopf, B. 2017. Elements of Causal Inference: Foundations and Learning Algorithms. MIT Press, Cambridge, Massachusetts.

43. Rubenstein, P. K., Weichwald, S., Bongers, S., Mooij, J. M., Janzing, D., Grosse-Wentrup, M., and Schölkopf, B. 2017. Causal consistency of structural equation models. Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence.

44. Rosch, E. 1978. Principles of categorization. In E. Rosch and B. B. Lloyd, editors, Cognition and Categorization: 27–48. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

45. Gärdenfors, P. 2000. Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge, Massachusetts.

46. Murphy, G. L. 2002. The Big Book of Concepts. MIT Press, Cambridge, Massachusetts.

47. Simon, H. A. 1962. The architecture of complexity. Proceedings of the American Philosophical Society 106(6): 467–482.

48. Anderson, P. W. 1972. More is different: Broken symmetry and the nature of the hierarchical structure of science. Science 177(4047): 393–396. DOI: 10.1126/science.177.4047.393.

49. Crutchfield, J. P., and Young, K. 1989. Inferring statistical complexity. Physical Review Letters 63(2): 105–108. DOI: 10.1103/PhysRevLett.63.105.

50. Shalizi, C. R., and Crutchfield, J. P. 2001. Computational mechanics: Pattern and prediction, structure and simplicity. Journal of Statistical Physics 104: 817–879. DOI: 10.1023/A:1010388907793.

51. Hoel, E. P., Albantakis, L., and Tononi, G. 2013. Quantifying causal emergence shows that macro can beat micro. Proceedings of the National Academy of Sciences of the United States of America 110(49): 19790–19795. DOI: 10.1073/pnas.1314922110.

52. Bengio, Y., Courville, A., and Vincent, P. 2013. Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(8): 1798–1828. DOI: 10.1109/TPAMI.2013.50.

53. Cohen, T. S., and Welling, M. 2016. Group equivariant convolutional networks. Proceedings of the 33rd International Conference on Machine Learning. Proceedings of Machine Learning Research 48: 2990–2999.

54. Bronstein, M. M., Bruna, J., Cohen, T., and Veličković, P. 2021. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. arXiv:2104.13478.

55. Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6): 716–723. DOI: 10.1109/TAC.1974.1100705.

56. Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6(2): 461–464. DOI: 10.1214/aos/1176344136.

57. Youvan, D. C. 2026. The Mathematics of Concept Formation: Discovery Calculus, Explanation Geometry, Theory Dynamics, and the Formal Birth of New Objects. ResearchGate. DOI: 10.13140/RG.2.2.23562.43209.

58. Youvan, D. C. 2026. Object Formation and Conceptual Resolution: Stable Quotients, Complete Refinement Lattices, and Functorial Observation. ResearchGate. DOI: 10.13140/RG.2.2.25921.72800.

59. Youvan, D. C. 2026. How Mathematics Makes Objects: Observation, Identity, and a New Algebra of Concept Formation. ResearchGate. DOI: 10.13140/RG.2.2.13338.81603.

60. Youvan, D. C. 2026. Local Observation and Global Objecthood: Sheaf Descent, Gluing Obstructions, and the Algebra of Distributed Concept Formation. Manuscript.